Université de Montréal

La faiblesse de volonté : conceptions classiques et dynamiques

Par
Jean-François Labonté

Département de philosophie
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures et postdoctorales
en vue de l'obtention du grade de Ph.D.
en philosophie

Septembre 2011

Université de Montréal
Faculté des études supérieures et postdoctorales

Cette thèse intitulée :

La faiblesse de volonté : conceptions classiques et dynamiques

Présentée par
Jean-François Labonté

a été évaluée par un jury composé des personnes suivantes :

Christine Tappolet

_____

président-rapporteur

Michel Seymour

_____

directeur de recherche

Daniel Laurier

_____

membre du jury

Ronald De Soussa

_____

examinateur externe

Christine Tappolet

_____

représentant du doyen

Résumé

La présente thèse expose, analyse et critique les positions classiques et modernes à l'égard de la nature et des causes de la faiblesse de volonté. L'identification du problème par Platon et Aristote a donné lieu à l'explicitation de principes et propositions portant sur la rationalité pratique en général et la motivation en particulier. Une discussion de ces principes et propositions est faite dans la mesure où ils ont conservé une certaine pertinence pour les théories modernes. Ce qui est devenu la conception standard de la stricte akrasie ainsi que son caractère prétendument paradoxal sont mis de l'avant.

Nous argumentons qu'une position sceptique à l'égard de la stricte akrasie ne peut pas reposer sur une version ou une autre de la théorie des préférences révélées et montrons qu'une description du processus décisionnel est nécessaire pour attribuer une préférence synthétique ou un meilleur jugement. Nous abordons le débat philosophique qui oppose une conception internaliste du lien entre le meilleur jugement et la décision à une conception externaliste, et soutenons, sur la base de résultats expérimentaux en psychologie cognitive et en neuroscience, que cette dernière conception est plus robuste, bien qu'imparfaite. Ces résultats ne vont pas toutefois à l'encontre de l'hypothèse que les agents sont des maximisateurs dans la satisfaction de leur préférence, laquelle hypothèse continue de justifier une forme de scepticisme à l'égard de la stricte akrasie. Nous exposons, par contre, des arguments solides à l'encontre de cette hypothèse et montrons pourquoi la maximisation n'est pas nécessairement requise pour le choix rationnel et que nous devons, par conséquent, réviser la conception standard de la stricte akrasie.

Nous discutons de l'influente théorie de Richard Holton sur la faiblesse de volonté non strictement akratique. Bien que compatible avec une conception non maximisante, sa théorie réduit trop les épisodes de faiblesse de volonté à des cas d'irrésolution.

Nous exposons finalement la théorie du choix intertemporel. Cette théorie est plus puissante parce qu'elle décrit et explique, à partir d'un même schème conceptuel, autant la stricte akrasie que l'akrasie tout court. Ce schème concerne les propriétés des distributions temporelles des conséquences des décisions akratiques et les attitudes prospectives qui

motivent les agents à les prendre. La structure de ces distributions, couplée à la dévaluation à l'égard du futur, permet également d'expliquer de manière simple et élégante pourquoi la faiblesse de volonté est irrationnelle. Nous discutons de l'hypothèse qu'une préférence temporelle pure est à la source d'une telle dévaluation et mentionnons quelques éléments critiques et hypothèses concurrentes plus conformes à une approche cognitiviste du problème.

Mots clefs : Philosophie – volonté – akrasie – décision – motivation – jugement – rationalité – maximisation – intertemporel

Abstract

This thesis explains, analyses and examines the classical and modern positions on the nature and causes of the weakness of will. Since Plato and Aristotle's identification of the problem, many principles and propositions on the subject of practical rationality in general and motivation in particular have been examined in details. These principles and propositions are being discussed on the basis that they are still somewhat relevant to modern theories. An emphasis is made on what is now known as the standard conception of strict akrasia and its supposedly paradoxical nature.

We argue that a skeptical position toward strict akrasia cannot be based on one version or another of the preference-revealed theory and we demonstrate that a description of the decision process is necessary to assign an overall preference or a better judgment. We discuss the philosophical debate on internalist and externalist conceptions of the connection between better judgment and decision. We then argue that, based on experimental results in cognitive psychology and neuroscience, the externalist conception, although imperfect, is stronger. But these experimental results are not incompatible with the hypothesis that agents are maximizers when it comes to the satisfaction of their preference. This hypothesis continues to justify a form of skepticism toward strict akrasia. However, we strongly argue against this hypothesis and we demonstrate why maximization is not absolutely necessary to rational choice; therefore, we have to revise the standard conception of strict akrasia.

We then discuss Richard Holton's influential theory on non-strictly akratic weakness of will. Although compatible with a non-maximizing conception, Holton's theory tends to reduce episodes of weakness of will to irresolution cases.

Lastly, we introduce the theory of intertemporal choice, a more potent theory that describes and explains, with the same conceptual schema, both strict and non-strict akrasia. This schema concerns the properties of temporal distribution of akratic decision's consequences and the prospective attitudes that motivate agents to make those decisions. Also, the structure of these distributions, along with the devaluation of the future, allows us to explain, clearly and simply, why weakness of will is irrational. We discuss the hypothesis

that this devaluation of the future is due to a pure temporal preference and we mention a number of critical elements and rival hypothesis more in keeping with a cognitive approach to the problem.

Keywords: Philosophy – will – akrasia – decision – motivation – judgment – rationality – maximization – intertemporal

Table des matières

Liste des figures

Pour Suzanne, mes parents, mes amis et mes collègues,
qui m'ont soutenu dans mon cheminement intellectuel.

Remerciements

Je n'aurais jamais pu mener ce projet doctoral à terme sans le soutien de mon directeur, de mes proches et de mes collègues de travail. Je tiens ici à leur témoigner ma reconnaissance.

Je veux d'abord souligner les bons conseils, les critiques constructives, la rigueur, la disponibilité et la curiosité de mon directeur de recherche, Michel Seymour. Bien que Michel ne soit pas un spécialiste de la question de la faiblesse de volonté, je l'ai choisi pour ses qualités intellectuelles et humaines. Il a accepté de me diriger de nouveau en dépit du fait que j'avais préalablement abandonné un projet de thèse antérieur.

Je ne peux manquer d'exprimer ma gratitude envers mes collègues (et amis) de travail, professeurs de philosophie, de sociologie et d'économie. Certains m'ont intellectuellement nourri en discutant avec moi du contenu d'articles et de bouquins. D'autres m'ont encouragé et certains m'ont divertie dans les moments cruciaux. Aussi, je remercie plus particulièrement Pierre-Alexandre Morneau-Caron, Pierre Blackburn, Tuan Nguyen, Thierry Toutant, Julien Lachapelle, David Meilleur, Kavin Hébert, Vincent-Pierre Martineau, Mathieu Lavoie, Véronique Grenier, Philippe Langlois, Steve McKay, et le Département de philosophie du Cégep de Sherbrooke pour leur soutien.

Je remercie également le Cégep de Sherbrooke dont le Service des ressources humaines a bien voulu m'octroyer un congé avec traitement différé à la session d'hiver 2011 pour la rédaction de la thèse.

Je ne peux manquer de témoigner mon immense gratitude envers Nadia Deslauriers, qui, par ses bons conseils linguistiques, me permet limiter les maux de têtes de mes évaluateurs.

Il est difficile de trouver des mots pour exprimer ma reconnaissance à l'égard de ma femme, Suzanne Poitras. Sa participation indirecte à la réalisation de la thèse ne s'est pas résumée à un soutien indéfectible et des encouragements répétés. Elle a due renoncer à un certain nombre de choses qui comptaient beaucoup à ses yeux (et aux miens!) comme des vacances en amoureux et des moments intimes libres de soucis intellectuels. Heureusement,

je compte bien rattraper le temps perdu avec la personne qui occupe la place la plus importante dans ma vie!

Avertissement

Pour les références à la littérature anglophone, j'ai utilisé les traductions françaises des articles et œuvres cités lorsque celles-ci étaient disponibles et me semblait adéquates.

J'utilise abondamment le terme « alternative » dans des contextes linguistique où il s'agit pourtant d'une faute sémantique et d'un anglicisme. En français, « alternative » dénote une situation qui comporte deux, et uniquement deux, possibilités entre lesquelles on doit choisir. Or, j'ai fait un usage du terme identique à son usage en anglais, c'est-à-dire qu'il dénote des situations qui présentent plusieurs alternatives, et pas seulement deux. Je pense qu'il est cependant approprié de le faire parce que les termes « options » et « choix », plus conformes à l'usage codifié, ne correspondent pas à certains aspects sémantiques importants du terme anglais. Contrairement au terme anglais, le terme « options » n'implique pas nécessairement le caractère mutuellement exclusif des possibilités, et le terme « choix » est souvent utilisé pour dénoter l'ensemble d'un processus décisionnel. Aussi, le terme « alternatives » me semble plus précis et être moins une source de confusions que les deux autres termes.

1
Introduction

Les choses n'iraient pas mieux si les hommes obtenaient justement ce qu'ils voulaient.

– Héraclite, Fragments

1.1 Le problème de la faiblesse de volonté ou l'art d'être son pire ennemi

Toutes sortes d'événements affectent notre niveau de bien-être tout au long de notre existence. Certains sont le résultat de processus naturels aveugles, alors que d'autres sont le résultat de processus décisionnels plus ou moins conscients de la part d'autrui. Dans chacune de ces possibilités, les événements et les facteurs en cause sont réalisés à l'extérieur de notre esprit et ce sont des choses sur lesquelles nous estimons souvent avoir peu de contrôle. En fonction de la nature de ces processus, notre réaction ira du blâme sévère – si l'on estime avoir été lésé dans nos droits – à la résignation spontanée – si l'on estime avoir été victime d'une simple malchance. Or, la totalité des événements qui affectent notre bien-être ne se réduit pas à ceux qui ont pour origine ces types d'événements et de facteurs. Il y a également toutes sortes de décisions que nous prenons et toutes sortes d'attitudes que nous entretenons qui altèrent pourtant, et de manière non négligeable, nos conditions de vie.

Que nous soyons à l'occasion l'artisan de notre propre malheur n'est pas un fait *prima facie* énigmatique. Je peux boire un verre de détergent reposant sur le comptoir, croyant à tort qu'il s'agissait d'une boisson aux fruits, et me retrouver à l'hôpital, ou je peux trébucher sur une pierre, que je n'avais pas vue, et me retrouver sur le sol avec un genou endolori. Dans la première situation, je chercherai à blâmer un responsable si j'en trouve un, alors que, dans la seconde, je finirai par me résigner à avoir un genou abîmé sans trop rouspéter. Ce qui est important ici réside dans le fait qu'aucune de ces décisions « inadéquates » n'a été prise dans le but de me nuire. Si je me suis nui, c'est que ces décisions n'ont pas été prises à la lumière de toutes les informations pertinentes. Si j'avais eu ces informations (il y a une

pierre dans ma trajectoire; il y a du détergent dans ce verre), j'aurais certainement agi autrement. Ces faits sont plutôt anodins et ne nécessitent pas d'explications particulières[1].

Par contre, il existe une classe de décisions que les agents prennent et d'attitudes qu'ils entretiennent *sachant* que cela affectera négativement leurs propres conditions de vie. On reporte les tâches ménagères, une visite chez le dentiste, la réalisation de travaux scolaires, le ramassage des feuilles mortes à l'automne, la prise d'un rendez-vous pour une mammographie, la lecture de rapports, la prise de médicaments, l'achat de cadeau jusqu'à la dernière minute, la rédaction d'une déclaration de revenus, une rupture amoureuse, etc. De même, on devance toutes sortes de décisions alors qu'on sait pertinemment que cela n'est pas à notre avantage de le faire. On devance une sortie avec des amis ou le visionnement d'un film alors qu'on doit s'acquitter de tâches plus importantes, l'achat d'un bien électronique, la consommation d'une sucrerie ou d'un dessert, une relation sexuelle alors qu'on n'a pas pris le temps de se protéger, etc. Si certaines de ces décisions ont un impact plutôt limité sur la qualité de notre bien-être général, d'autres ont un impact beaucoup plus marqué. Une quantité considérable de personnes perdent leur emploi, leur famille, ou altèrent sérieusement leur santé parce qu'ils n'ont pu s'empêcher de prendre, sur une base régulière, de mauvaises décisions en toute connaissance de cause – qu'on pense aux toxicomanes et aux alcooliques, aux fumeurs, aux dilettantes chroniques, aux joueurs compulsifs, pour ne nommer que les cas les plus patents.

On dispose d'une panoplie de termes pour décrire ce genre de décision. Suivant les cas, on parle de compulsivité, d'impulsivité, de dépendance, d'impatience ou de tentation, mais aussi de procrastination, de non-observance, d'irrésolution ou d'apathie. Le meilleur terme *générique* qu'on peut utiliser pour décrire ces phénomènes en apparence disparates a été forgé initialement par la tradition philosophique. Ce genre de décision relèverait de cas de *faiblesse de volonté*. L'appellation semble de prime abord un peu inadéquate parce que le substantif « faiblesse » n'apparait pas approprié pour décrire des cas d'impulsivité ou de

[1] Une précision s'impose ici. Ce n'est pas parce que des faits de ce genre ne nécessitent pas d'explications qu'ils ne suscitent pas de problèmes philosophiques. Pour déterminer, par exemple, si une décision est rationnelle, on doit tenir compte des informations dont disposent les agents et non seulement des informations *disponibles* dans la structure de l'environnement. On peut mettre l'accent sur le *point de vue de l'agent* dans la recherche d'un critère de rationalité. Ou, *a contrario*, estimer que le point de vue de l'agent est insuffisant pour déterminer si une décision est rationnelle ou non. Car, si l'on ne s'appuie que sur ces facteurs internes, il devient très difficile de condamner des décisions comme étant irrationnelles alors qu'elles sont vouées à l'échec (Ogien, 2003).

compulsivité où les personnes semblent vouloir avec *trop de force* quelque chose. Mais cela n'est pas justifié parce que si une personne veut trop quelque chose, alors elle ne veut pas assez quelque chose d'autre. Pour tous les cas mentionnés plus haut, il semble y avoir en amont une sorte de déficit motivationnel.

Bien que le concept de faiblesse de volonté ait acquis aujourd'hui un contenu très technique – tant en philosophie qu'en psychologie ou en économie –, le phénomène qu'il est censé décrire est relativement facile à identifier à partir d'une conception intuitive. Les décisions prises lors d'épisodes de faiblesse de volonté ne sont pas motivées par de bonnes raisons ou par des raisons qui sont suffisamment fortes, et produisent, par le fait même, des résultats qui vont systématiquement à l'encontre de l'intérêt de ces personnes. Ces choix font d'ailleurs régulièrement l'objet de critiques de la part d'autrui, mais également de regrets subséquents de la part des personnes qui les font. Or, la conjugaison de tous ces éléments fait de la faiblesse de volonté le paradigme de l'irrationalité pratique et pose par conséquent un certain nombre de défis pratiques et théoriques importants.

1.1.1 Les défis pratiques : les stratégies de contrôle de soi

Étant donné la prévalence des cas de faiblesse de volonté et leurs conséquences parfois importantes, les défis pratiques que pose ce paradigme de l'irrationalité sont aussi sociaux qu'individuels. Le problème que représente la faiblesse de volonté est extrêmement prévalent. Il touche toutes les couches de la population – mais à divers degrés – et est à la racine de plusieurs problèmes sociaux d'envergure comme les problèmes de santé publique, les grossesses non désirées et la non-observance des ordonnances médicales, le non-respect des normes morales comme celles proscrivant le harcèlement sexuel (Ariely, 2008), les problèmes d'éducation, comme la non-assiduité scolaire et le décrochage (Pychyl, Lee, Thibodeau & Blunt, 2000), et les problèmes d'ordre économique, comme la quasi-inexistence de l'épargne des ménages (Akerlof, 1991) et la pauvreté récalcitrante (Heath, 2009). D'ailleurs, 15 % à 20 % des adultes se disent affectés d'un problème de procrastination chronique et près de 95 % d'entre eux considèrent la procrastination mauvaise et souhaiteraient la réduire (Steel, 2007).

Or, l'élaboration de stratégies de contrôle de soi est le principal défi pratique que pose la faiblesse de volonté. Les stratégies les plus adéquates ne sont pas nécessairement de nature

individuelle. Les dispositifs institutionnels peuvent être mis en place pour pallier les défauts de volonté des citoyens dans certaines sphères d'activité bien précise. Par exemple, les mesures encadrant la vente et la distribution d'alcool, les règles de délai obligatoire minimal pour se marier ou pour faire l'acquisition d'armes meurtrières sont des facteurs de « refroidissement » qui limitent la surconsommation et l'impulsivité. D'ailleurs, les destinataires des mesures visant un contrôle accru de soi ne sont pas non plus nécessairement des personnes prises individuellement. Il peut s'agir de groupes sociaux plus ou moins importants ou même des nations entières (Elster, 1986, 2007). Les dispositifs constitutionnels, par exemple les règles encadrant la politique monétaire d'un pays, et la délégation de sa responsabilité aux banques centrales, permettent d'éviter qu'elle soit utilisée à des fins partisanes pour la poursuite exclusive d'intérêts à court terme.

En fait, lorsqu'on parle de stratégies de contrôle de soi, on doit garder à l'esprit qu'elles ne sont pas toutes au même niveau opérationnel et qu'elles peuvent former une sorte de mille-feuille motivationnel. Les stratégies vont des mesures psychologiques directes (ex. : faire un effort pour maintenir son attention fixée sur son objectif, se rappeler pourquoi on accomplit telle besogne déplaisante, redécrire ses options pour modifier leur degré d'attraction ou de répulsion, etc.) aux mesures environnementales (ex. : vider ses bouteilles d'alcool dans les toilettes, étudier à la bibliothèque, arracher le bouton « *snooze* » de son réveille-matin, etc.) et sociales les plus indirectes (ex. : promettre à quelqu'un qu'on arrêtera de fumer, se retrouver dans une situation où l'on est encadré par des parents, établir une loi limitant la quantité de machines à sous dans les quartiers défavorisés, etc.). On peut faire un usage individuel et collectif d'une série de stratégies différentes, situées à divers niveaux, pour résoudre un même problème de faiblesse de volonté. Le choix des stratégies doit être commandé par leur efficacité et leur coût d'implantation, et par l'ampleur du problème qu'on vise à résoudre.

En fait, une bonne partie de la littérature psychologique – tant scientifique que populaire – qui porte d'une manière ou d'une autre sur la faiblesse de volonté aborde le problème à partir d'une perspective purement pratique. On établit un diagnostic sur les causes probables et facteurs corrélés, et l'on établit ensuite des stratégies directes de renforcement de la

volonté ou des plans globaux de modification du caractère[2].

D'une manière générale, les stratégies de contrôle de soi élaborées par les psychologues cliniques, ainsi que les personnes qui vivent sur une base régulière des problèmes de faiblesse de volonté, s'appuient souvent sur des conceptions intuitives de ses cas de figure. Les personnes qui vivent ces problèmes décisionnels mettent en place – souvent spontanément – des stratégies relativement efficaces sur la base d'une connaissance somme toute assez limitée de la nature du problème. Si ce n'est pas suffisant, ils font, dans certains cas, appel à des psychologues professionnels pour forger des stratégies plus adaptées, qui tiendront compte des traits caractériels de leurs patients. Certaines des stratégies mises en place sont efficaces, alors que d'autres apparaissent clairement douteuses. Dans un ouvrage de psychologie populaire sur la procrastination, on suggère, par exemple, aux personnes de devenir pour elles-mêmes leur plus grand fan (Burka & Yuen, 1983). D'ailleurs, à défaut d'une théorie détaillée, la littérature de psychologie clinique et populaire a tendance à proposer toujours les mêmes recettes et à ignorer les stratégies environnementales et sociales.

Ce qui manque à l'élaboration des stratégies de contrôle de soi efficaces et économiques est un arrière-plan conceptuel et théorique détaillé, et suffisamment robuste, de la faiblesse de volonté. Un tel arrière-plan devrait fournir des indications précieuses pour l'élaboration de stratégies efficaces, tant pour les personnes qui éprouvent des problèmes motivationnels que pour les psychologues cliniques, les législateurs et professionnels qui participent à des programmes d'ingénierie sociale. Cette tâche incombe à la psychologie cognitive, à la neurologie et à l'économie comportementale, mais aussi à la philosophie, qui a une longue tradition dans l'examen du phénomène et de sa conceptualisation.

[2] Parmi la liste pour le moins très hétérogène des causes les plus communes et les facteurs corrélés les plus forts de la faiblesse de volonté ont trouve l'âge et le sexe, l'intoxication par une substance addictive, l'anxiété d'évaluation, le refus d'autorité, le caractère attrayant ou repoussant des tâches à effectuer, la distance temporelle des récompenses et punitions, les traits de caractère comme l'estime de soi et la dépression, l'ouverture aux nouvelles expériences, l'incertitude des résultats, l'implémentation d'habitude, le perfectionnisme et même des facteurs génétiques (Steel, 2007).

1.1.2 Les défis théoriques de la faiblesse de volonté : sa nature, ses formes et son caractère paradoxal

Avec une conception intuitive, on peut diagnostiquer la plupart des cas de faiblesse de volonté. Par contre, lorsqu'on essaie de raffiner le concept et de spécifier le contenu qu'il devrait avoir dans le but d'éviter de rechercher à l'aveugle des solutions ou de faire du pilotage à vue dans la gestion de notre motivation, beaucoup de questions et de problèmes émergent. Les solutions apportées à ces problèmes n'ont toutefois pas toutes une incidence pratique. Certaines sont d'ordre conceptuel, d'autres spécifient les constituants ultimes de la motivation, alors que d'autres encore portent sur les procédures décisionnelles et les normes de la rationalité pratique. Mais l'identification des problèmes et des solutions acceptables contribue à notre compréhension générale du phénomène et doit servir d'arrière-plan théorique et conceptuel pour l'élaboration de solutions adéquates, de la même manière que les théories et concepts en biologie fondamentale servent à l'élaboration de médicaments, de mesures de prévention et d'interventions médicales.

Une enquête sur quelque phénomène que ce soit ne commence pas à partir de rien (*from scratch*), et une enquête sur la nature de la faiblesse de volonté ne fait pas exception à la règle. Les théoriciens abordent le phénomène avec initialement une conception ou une théorie de base de la motivation, des normes de rationalité ou des procédures décisionnelles typiques des agents rationnels, voire des agents tout court. Même les toutes premières tentatives d'explication, qui remontent à Platon et Aristote, incluent des considérations générales explicites sur des normes jugées constitutives de la rationalité et de la manière dont les agents prennent des décisions. En fait, les critiques des tentatives théoriques de conceptualisation et d'explication de la faiblesse de volonté portent la plupart du temps sur la manière dont les théoriciens conçoivent *en général* les facteurs, processus et normes décisionnels.

D'ailleurs, le principal défi que pose la faiblesse de volonté est de la concilier avec la manière dont les théoriciens conçoivent habituellement la motivation et les processus menant aux décisions. Lorsque les théoriciens proposent des descriptions (même sommaires) des processus décisionnels des agents rationnels, la faiblesse de volonté acquiert un caractère si *paradoxal* que nombres d'entre eux en viennent à penser que,

malgré les apparences, elle n'existe probablement pas. Il s'agit d'un problème particulièrement difficile à résoudre et qui a fait l'objet d'âpres débats depuis l'Antiquité.

Les sources du paradoxe sont multiples. Mais la principale vient du fait que la plupart des théories du choix rationnel estiment que les agents sont des *maximisateurs* (de bien-être ou de satisfaction de leurs préférences). Aussi, les épisodes de faiblesse de volonté sont énigmatiques dans la mesure ou nous souscrivons de manière générale à l'idée que les agents tendent systématiquement, en prenant des décisions sur la base d'informations dont ils disposent, d'améliorer, quand cela est possible, leurs conditions d'existence ou du moins d'en limiter une dégradation inévitable. Or, quand nous vivons un épisode de faiblesse de volonté, nous nous nuisons, semble-t-il, en toute connaissance de cause ou nous négligeons volontairement l'examen des conséquences des choix que nous envisageons de faire. C'est pourquoi, bien que ces épisodes soient si communs et à certains égards anodins, la tradition philosophique – et dans une certaine mesure économique – a considéré, à juste titre, la faiblesse de volonté comme étant le paradigme de l'irrationalité pratique, et à cet égard, le phénomène a reçu une attention non négligeable.

Une façon de dissoudre en partie le caractère paradoxal du phénomène est de distinguer deux formes de faiblesse de volonté. Il y a la faiblesse de volonté au sens large et au sens étroit. Au sens large, la faiblesse de volonté consiste en décisions que les agents prennent à l'encontre de leur propre intérêt parce que leurs préférences, leurs estimations de leur propre bonheur futur et même dans certains cas leurs croyances – en particulier dans les cas d'acquisition de croyances motivées – sont instables. On parlera alors de faiblesse de volonté diachronique, aveugle, ou d'akrasie tout court. Tandis qu'au sens étroit, la faiblesse de volonté concerne les décisions qu'on prend, semble-t-il, à l'encontre de son intérêt, *mais* en toute connaissance de cause, et en estimant que d'autres options sont supérieures. On parlera alors de faiblesse de volonté synchronique, les « yeux ouverts », ou de stricte akrasie.

Prima facie, les deux formes sont paradoxales parce qu'elles vont toutes deux à l'encontre de l'hypothèse de la maximisation. Mais, en fait, la première forme n'est pas plus paradoxale que le sont les mauvaises décisions et les erreurs en général. Comme aucun agent n'est omniscient, même dans l'erreur il peut demeurer un maximisateur. Dans les cas où il n'y a manifestement pas d'erreur, mais un changement dans les préférences (ou les

goûts), les cas de faiblesse de volonté seront vues comme la maximisation de la satisfaction des préférences *du moment*.

La tradition philosophique ne s'est que peu ou pas du tout intéressée à cette forme. C'est la seconde forme qui a suscité d'âpres débats qui perdurent depuis l'Antiquité. Dans le cas de la stricte akrasie, il est beaucoup plus difficile de dissoudre le paradoxe. On ne peut pas dire que l'agent strictement akratique ignore certains faits pertinents ou que ses préférences sont instables.

Le défi principal que la faiblesse de volonté a posé pour la tradition philosophique est celui de l'existence putative de la stricte akrasie. Aussi, la plus grande portion de la littérature consacrée au sujet porte directement sur ce problème, et plus précisément sur l'existence de la stricte akrasie *morale*. En s'attaquant à ce défi, les philosophes, mais également les chercheurs en sciences cognitives et en économie, ont élaboré des modèles – quelques fois simples, quelques fois complexes – des processus décisionnels. Les modèles proposés sont censés appuyer les positions de chacun à l'égard du caractère paradoxal de la forme synchronique de la faiblesse de volonté. Aussi, le débat sur l'existence de la stricte akrasie se transforme souvent en débat sur les éléments motivationnels de base et dérivés, sur les procédures décisionnelles, sur les normes d'évaluation des choix ou les normes d'interprétation de ceux-ci.

Étrangement, la description et l'explication de l'akrasie tout court, ou faiblesse de volonté diachronique, sont, dans la littérature philosophique, souvent délaissées. Comme son caractère paradoxal est plus facile à dissoudre, on estime plus stimulant de s'attaquer à la stricte akrasie. Pourtant, l'examen de ces cas est tout aussi intéressant et, qui plus est, fournit des éléments importants pour comprendre les cas putatifs de stricte akrasie. La manière dont les agents choisissent à l'encontre de leurs intérêts exemplifie des propriétés communes à toutes les figures de la faiblesse de volonté. Or, on ne doit pas seulement décrire les motivations qui poussent les agents à prendre des décisions allant à l'encontre de leurs intérêts, mais également décrire les propriétés des décisions elles-mêmes. D'ailleurs, en tenant compte des propriétés des décisions akratiques, le théoricien se met en meilleure position pour évaluer des stratégies de contrôle de soi ou formuler des indications utiles pour leur élaboration.

En dépit de leur importance théorique, la littérature philosophique ne s'est pas beaucoup intéressée à la question de la nature des décisions akratiques. Aussi, on doit se rabattre sur la psychologie théorique et les sciences cognitives pour obtenir des éléments de réponse.

1.2 Positions défendues et plan du texte

L'essentiel de mon propos est critique, mais comporte également des thèses et des hypothèses substantielles sur la nature des deux formes de faiblesse de volonté. Aussi, je m'attaquerai en particulier au débat sur la stricte akrasie et ses sources.

J'exposerai d'abord (**chapitre 2**) les premières tentatives d'identification et d'explication du phénomène de la faiblesse de volonté, en particulier celles qu'on trouve chez Platon et Aristote. L'examen des dialogues socratiques révèle un certain nombre de propositions théoriques qui, sans nécessairement faire école, ont eu une longue vie. Platon a défendu une forme de cognitivisme motivationnel qui n'a pas aujourd'hui beaucoup d'adhérents, mais qui étrangement semble avoir conditionné ce qui est devenu la conception standard de la stricte akrasie, à savoir le fait d'agir à l'encontre de son meilleur jugement. Il a également exposé, sans lui donner cependant de nom, une conception maximisante du bien-être comme norme constitutive du processus décisionnel. En conjonction avec son cognitivisme motivationnel, la norme de maximisation constitutive justifie sa position sceptique à l'égard de la stricte akrasie. L'explication platonicienne de l'akrasie tout court réside quant à elle dans la difficulté de choisir en conformité avec une norme de proportionnalité du plaisir anticipé et de l'anticipation du plaisir. Du côté d'Aristote, les explications sont plus variées et ne sont donc pas unifiées en une seule théorie. Elles vont de l'oubli d'informations pertinentes, ou d'attitudes négligentes, à la contamination de la délibération rationnelle par des émotions. Cette dernière explication est particulièrement intéressante parce qu'elle est fondée sur l'idée répandue que les émotions figurent parmi les principaux facteurs d'irrationalité pratique. Or, il s'agit d'une hypothèse très contestable. À la lumière des résultats expérimentaux en neurologie du choix, il semblerait même qu'un déficit émotionnel expliquerait certains cas pathologiques de comportements impulsifs. J'exposerai également brièvement le volontarisme médiéval qui, pour des raisons essentiellement théologiques, a détaché et isolé les décisions de ses déterminants motivationnels autant cognitifs que désidératifs et affectifs. Cela a impliqué la postulation

d'une faculté mentale volitive indépendante, laquelle est une hypothèse douteuse, mais a également conduit la tradition médiévale à voir dans la faiblesse de volonté un problème moteur ou un problème d'exécution de décision, et non pas un problème motivationnel.

Après une discussion des positions de Platon, d'Aristote et du volontarisme médiéval, j'exposerai (**chapitre 3**) ce qui est devenu aujourd'hui la conception standard de la stricte akrasie. Je ferai quelques amendements que j'estime essentiels, notamment l'idée de faire porter la conception sur des décisions plutôt que sur des actions, et l'idée de décrire l'attitude des agents comme des estimations plutôt que comme des jugements. La référence aux décisions est plus appropriée parce que plus conforme à nos intuitions. Tandis que la référence aux estimations est conceptuellement plus neutre parce qu'elle s'accorde autant avec les conceptions désidératives (et émotives) de la motivation qu'avec la conception cognitive.

Avec la conception Standard comme référence, j'examinerai ensuite (**chapitre 4**) ce que j'estime être une des positions sceptiques les plus fortes à l'égard de la stricte akrasie, à savoir la théorie des préférences révélées. Selon cette théorie, les décisions des agents sont censées être suffisamment transparentes pour que nous puissions en dériver les préférences qui les motivent sans que nous ayons à faire des hypothèses hasardeuses sur des processus internes menant à la production de ces décisions. En conjonction avec l'hypothèse de la maximisation, la théorie des préférences révélées implique l'inexistence de la stricte akrasie. Or, je montrerai qu'il y a un nombre important de raisons qui nous permettent de croire qu'on ne peut faire l'économie d'hypothèses sur ces processus internes. J'arrive à la conclusion que la théorie des préférences révélées est erronée et que, par conséquent, il est inapproprié de rejeter l'existence de cas de stricte akrasie sur cette base.

Si nous ne pouvons faire l'économie d'une théorie ou d'un modèle descriptif du processus décisionnels, la question à se poser alors est quel processus est typique de la stricte akrasie? J'offrirai (**chapitre 5**) des éléments de réponse à cette question, qui permettront d'arrimer les deux conceptions traditionnelles de la motivation (cognitive et désidérative ou affective) et la conception standard. Toujours en conformité avec l'hypothèse de la maximisation, je montrerai comment les agents forment au moyen de la délibération (consciente ou inconsciente) un meilleur jugement ou une préférence synthétique complète à l'égard des choix qui s'offrent à eux. Je montrerai que, avec une description détaillée des étapes du

processus décisionnel, on évite de confondre des cas authentiques de stricte akrasie avec d'autres types de décisions irrationnelles et de problèmes motivationnels, même s'ils semblent satisfaire les conditions énoncées par la conception standard. Aussi, une description détaillée du processus permet d'isoler avec plus de précision le problème motivationnel putatif de l'agent strictement akratique et l'objet du débat moderne sur l'existence de la stricte akrasie.

J'examinerai ensuite (**chapitre 6**) le débat entre internalistes et externalistes portant sur la possibilité d'échec dans l'application du résultat de la délibération, le meilleur jugement de l'agent. Je présenterai l'internalisme fort de Richard Hare – la position classique dans ce débat – l'externalisme d'Alfred Mele, ainsi que l'internalisme faible de Donald Davidson. Je mettrai en relief certaines difficultés qu'on peut isoler dans les termes mêmes du débat, et surtout sur la notion de jugement inconditionnel, qui est au cœur de l'approche davidsonienne et qui est incontournable dans les débats actuels. Je mentionnerai finalement certains résultats expérimentaux, notamment en neurologie, qui militent manifestement en faveur de l'hypothèse externaliste. Ces résultats mettent en relief le fait que les agents sont pourvus d'un système motivationnel distinct de leur système d'évaluation consciente des alternatives de choix. Ces résultats sont suffisamment probants pour ébranler sérieusement un scepticisme s'appuyant sur la conception internaliste de la motivation.

Je m'attaquerai ensuite (**chapitre 7**) à l'hypothèse de la maximisation, qui est sans doute le pilier de la position sceptique. Il s'agit d'une hypothèse qui semble particulièrement robuste parce qu'elle peut se décliner en plusieurs variantes et porter sur des éléments motivationnels de notre choix (considérations cognitives, désidératives ou affectives) ou avoir une forme interprétative « *comme si* ». En fait, j'entends montrer que la maximisation est une hypothèse erronée parce qu'inapplicable dans la plupart des contextes réels de choix. Les agents rationnels utilisent toutes sortes de règles décisionnelles qui les amènent à négliger des informations pertinentes pour la délibération, mais leur offrent des perspectives de réussite plus attrayantes. À l'aide d'exemples documentés par des recherches sur les heuristiques décisionnelles, je mettrai en relief le fait que très peu de décisions, voire presque aucune, ne sont motivées par un meilleur jugement ou par une préférence synthétique complète. En plus de dissoudre une partie du caractère paradoxal de la stricte akrasie, rejeter l'hypothèse de maximisation a des implications importantes pour

la conception standard. Elle implique notamment que la conception standard produit un sous-diagnostic des cas de stricte akrasie.

Je me tournerai ensuite (**chapitre 8**) vers la théorie de la faiblesse de volonté de Richard Holton. Jouissant actuellement d'une forte popularité, sa théorie ne vise pas à résoudre le débat sur l'existence de la stricte akrasie. Holton s'intéresse plutôt à l'akrasie tout court, qu'il décrit comme une forme irrationnelle de révision de plans et de résolutions. En plus de déplacer le centre d'attention vers l'autre forme de faiblesse de volonté, Holton propose une théorie qui ne repose pas sur l'hypothèse de la maximisation et qui est tout à fait compatible avec toute une variété de procédures et de règles décisionnelles. Sa difficulté cependant vient du fait qu'elle n'offre pas vraiment un modèle complet pour décrire, expliquer et comprendre la dynamique motivationnelle des agents, et réduit l'akrasie à la classe trop limitée des cas d'irrésolution.

J'exposerai finalement (**chapitre 9 et 10**) la conception inter-temporelle du choix. Cette conception offre, je pense, les meilleurs éléments de réponse au problème de la faiblesse de volonté. Elle s'appuie sur l'idée que les agents ne sont pas indifférents à la manière dont sont temporellement distribuées les conséquences attendues des alternatives de choix et choisissent en fonction de ces distributions. Loin d'être exotique, cette attitude explique toutes sortes de décisions et de comportements autant chez les agents humains que chez les organismes végétaux. Or, elle expliquerait également les épisodes de faiblesse de volonté. Les grands avantages de la conception inter-temporelle sont qu'elle nous permet, d'une part, d'unifier d'une manière simple et élégante l'explication de la stricte akrasie et de l'akrasie tout court et, d'autre part, de fournir une explication substantielle (et non simplement formelle) de l'irrationalité de la faiblesse de volonté. J'exposerai la conception inter-temporelle en deux volets. Le premier (**chapitre 9**) porte sur les propriétés des alternatives de choix en concurrence et sur les décisions akratiques elles-mêmes. Ce volet fournit des éléments d'explication du caractère irrationnel de la faiblesse de volonté. Le second (**chapitre 10**) porte sur les propriétés des attitudes prospectives qui motivent le choix d'une alternative typiquement akratique. L'identification des attitudes responsables chez l'agent akratique de sa dévaluation du futur a fait l'objet d'une littérature abondante, notamment en psychologie du choix. Beaucoup de chercheurs estiment qu'il y a une préférence temporelle au sein de l'espace motivationnel de l'agent qui expliquerait une telle

dévaluation. Mais il n'y a pas de consensus sur le fait de savoir s'il s'agit d'une préférence pure ou dérivée, ou si la dévaluation du futur ne serait pas plutôt attribuable à un problème cognitif de représentation des événements futurs. Comme les débats sur le sujet sont récents, j'exposerai ces points de discorde sans prendre position. Je mentionnerai cependant une conséquence importante que l'hypothèse de la dévaluation du futur semble avoir pour le débat philosophique, celle d'inverser le problème. Que la faiblesse de volonté soit le résultat d'une préférence temporelle pure ou d'une myopie cognitive, il appert que c'est moins son occurrence qui requiert une explication que son absence.

1.3 Soucis théoriques, empiriques et méthodologiques

L'utilité d'une investigation théorique du phénomène de la faiblesse de volonté pour l'élaboration de solutions ne doit évidemment pas impliquer un dédain pour les intuitions communes sur le sujet. À défaut d'avoir une théorie constituée, suffisamment robuste et pleinement opérationnelle, nos intuitions sont souvent de bons points de départ pour entreprendre une investigation et tester nos hypothèses en cours de route. Cela n'implique pas pour autant qu'elles arriveront toutes intactes au point d'arrivée. Il est utile de suivre le principe qu'il faut, dans la mesure du possible, ménager nos intuitions lorsque vient le temps de traiter un problème aussi commun que celui des échecs motivationnels, mais, pour faire suite à une argumentation serrée, on peut arriver à la conclusion qu'on doit délester quelques-unes de ces intuitions. D'ailleurs, les récentes découvertes et résultats expérimentaux en neurologie, science cognitive, psychologie de la motivation, théorie de la décision et économie comportementale jettent une lumière nouvelle sur le phénomène, qui ne cadre pas toujours bien avec nos intuitions les plus fortes. Aussi, ma démarche sera en partie guidée par des intuitions, qui sont, jusqu'à preuve du contraire, suffisamment robustes, et par des résultats expérimentaux.

Les stratégies de contrôle de soi, que les personnes mettent en place pour eux-mêmes ou que les professionnels prescrivent, sont également de précieuses sources d'informations. Les stratégies qui obtiennent un taux de réussite important, de même que celles qui impliquent systématiquement un échec, nous aident à dresser un portrait des difficultés motivationnelles de l'agent akratique, même si elles sont élaborées de manière intuitive.

Aussi, je ferai référence en de multiples occasions à ces stratégies autant pour critiquer des positions que pour en appuyer.

Du point de vue théorique, ma démarche sera guidée par des modèles conceptuels des processus décisionnels réalistes et suffisamment riches pour rendre compte de détails signifiants. Mettre à profit des résultats expérimentaux pour sélectionner les bons modèles n'est pas une tâche aisée. Le phénomène de la faiblesse de volonté a été l'objet d'une quantité impressionnante d'articles dans diverses revues de psychologie théorique et d'économie comportementale. Aussi, il n'est pas toujours facile de s'y retrouver. Ceci vaut la peine d'être mentionné parce que, parmi les difficultés que le philosophe éprouve lorsqu'il aborde la littérature en psychologie, on trouve au premier chef l'absence d'une synthèse globale des dernières avancées – résultat d'un travail préalable de méta-analyse – qui prend la forme d'une théorisation des phénomènes (Elster, Lowenstein : 1992), et l'absence d'un idiome commun. La littérature psychologique propose souvent, à l'égard de la faiblesse de volonté, des explications moins « ambitieuses » que celles qu'on trouve dans la littérature philosophique et économique. À quelques exceptions près, les théoriciens qui mènent des études dans le domaine de la psychologie n'abordent pas le problème de la faiblesse de volonté dans « toute sa généralité », mais concentrent leurs efforts sur des cas de figure comme la procrastination et l'impatience, l'indécision et l'irrésolution, les caprices, l'imprudence et l'accidie, la non-observance, les compulsions et la dépendance. Les enquêtes empiriques dans lesquelles ils s'engagent visent la plupart du temps à identifier des facteurs corrélés avec ces diverses figures. De l'ensemble de ces recherches, il en résulte une mine d'informations dont la variété et la richesse n'ont d'égal que le manque de repères généraux. Aussi, il est plus facile pour le philosophe de se retrouver dans la littérature économique, malgré le fait que cette dernière fait une utilisation poussée d'équations mathématiques souvent absconses. En fait, le niveau différent de « fédération » des idiosyncrasies dans ces disciplines donne au philosophe l'impression que l'économie ressemble à un jardin français et la psychologie davantage à un jardin anglais. En effet, la psychologie théorique, et même l'économie comportementale, offre un ensemble de petites théories ou explications plus ou moins hétérogènes, mais pas de théories unifiées. Aussi, le travail de synthétisation, que doit faire le philosophe qui discute d'un problème qui est traité si différemment du point de vue théorique et empirique dans les disciplines

scientifiques, va nécessairement être incomplet et reposer quelques fois sur des interprétations approximatives. Mon travail théorique ne fait pas l'économie de ce genre de difficultés, mais essaye de les limiter.

En fait, un des défis théoriques importants dans l'explication de la nature de la faiblesse de volonté est justement de montrer que le phénomène n'est pas particulièrement exotique, mais est intelligible dans le cadre des meilleures théories *générales* de la décision et de la motivation. Sinon, on serait obligé de faire des hypothèses *ad hoc* qui entraîneraient une conception du phénomène peu robuste, ce qui s'avèrerait au final peu convaincant.

2
Les premières théories sur l'akrasie

Tout art et toute recherche*, de même que toute* action *et toute délibération réfléchie, tendent, semble-t-il, vers quelque bien. Aussi a-t-on eu parfaitement raison de définir le bien : ce à quoi on tend en toutes circonstances.*

– Aristote

2.1 L'intérêt théorique des conceptions de Platon et d'Aristote

Il est utile de débuter en faisant une brève excursion historique du côté des premières théories sur la faiblesse de volonté. Je vais examiner plus précisément les conceptions et les éléments théoriques qu'on trouve dans l'œuvre de Platon et d'Aristote. On peut estimer que, d'un point de vue strictement historique, les conceptions de ces philosophes occupent une place de choix parce qu'elles représentent les premières tentatives de poser le problème, de le conceptualiser et de l'expliquer. Mais j'estime pour ma part que l'intérêt pour les œuvres de Platon et Aristote ici va au-delà d'un intérêt pour l'étude des conceptions primitives. Sans dénigrer les enquêtes historiques et l'exégèse des textes comme des activités en soi, les conceptions de ces auteurs de l'Antiquité présentent encore, en dépit de leur caractère un peu suranné, un intérêt théorique indubitable pour quiconque étudie le phénomène de la faiblesse de volonté aujourd'hui. Platon et Aristote partagent des intuitions, des présupposés et des positions théoriques qu'on retrouve au cœur de certaines conceptions et théories sophistiquées en vogue actuellement, ou à l'égard desquels certaines théories et conceptions actuelles s'inscrivent en porte-à-faux. C'est dans cela que réside le principal intérêt théorique des conceptions platoniciennes et aristotéliciennes, et c'est la raison pour laquelle j'y consacre un chapitre. Mais je compte aussi mettre en relief certaines idées avancées par Platon et Aristote qui ont moins bien vieilli et mentionner quelques éléments critiques à leur égard.

Les exégètes voient dans l'œuvre de Platon, et en particulier dans le dialogue *Le Protagoras*, les premières discussions sur l'akrasie. Or, il ne revient pas seulement à Platon d'avoir été le premier philosophe écrivain à identifier le problème de la faiblesse de

volonté, même s'il revient à Aristote de l'avoir identifié conceptuellement – avec le terme « *akrasie* ». Il a orienté d'une certaine manière la discussion sur le phénomène de la motivation en général. Il a su dégager en partie les problèmes de philosophie de l'action du giron moral. Ce qui en fait à ce titre un des premiers théoriciens de l'action au sens moderne. Il a été capable de concevoir le problème de la faiblesse de volonté comme un problème de motivation qui dans certaines situations n'a pas grand-chose à voir avec la défaillance morale. Platon semble suggérer que les agents recherchent tantôt à maximiser leur plaisir, tantôt à s'orienter en fonction de l'idée de justice, mais sans fournir pour autant une théorie de leurs interactions possibles au sein de processus cognitifs de délibération. Mais ce qu'il est important de rappeler est l'orientation platonicienne vers une conception des problèmes motivationnels hors du domaine strictement moral. La philosophie de l'action a, au 20e siècle, achevé un renversement de l'ordre explicatif. On accepte aujourd'hui tout naturellement l'idée qu'il y a des motivations au sens large du terme, en plus des motivations proprement morales. Et que l'explication de ces dernières doit figurer comme un chapitre d'un ouvrage plus général pour ce qui motive les agents à prendre telle décision plutôt qu'une autre, à agir de telle façon plutôt qu'une autre[3].

Mais l'intérêt théorique le plus important que présente ici la théorie de Platon est qu'il a formulé en quelque sorte une proto-version de ce qui est devenu par la suite la doctrine de *l'agent maximisateur d'utilité espérée*. Élaborée au 20e siècle dans la discipline de la micro-économie, cette théorie complexe fondée sur des axiomes de cohérence formelle des préférences et des choix a joui – et jouit toujours – d'une grande popularité parmi les théoriciens des sciences sociales. Si cette théorie est correcte, cela signifie qu'il n'y a pas, à proprement parler, de décisions *strictement* akratiques ou d'épisodes de faiblesse de volonté synchronique. Beaucoup du scepticisme contemporain à l'égard de la possibilité de la stricte akrasie est tributaire du moins implicitement de cette conception de l'agent rationnel. Pour ce qui est de la faiblesse de volonté en général, Platon en identifie la cause dans une

[3] Par contre, je ne m'explique pas le peu d'intérêt qu'ont les philosophes actuels pour le problème de la motivation pour des agents non humains. Si Platon a eu raison de traiter à l'occasion l'akrasie comme un problème de motivation humaine en général, l'autre pas consisterait à examiner du moins la possibilité qu'il y ait de la faiblesse de volonté chez les animaux. Or, il n'y a pas, à ce que je sache, de considérations de ce genre dans la littérature philosophique. Pourtant, comme je vais le montrer au chapitre 9, des chercheurs en psychologie ont mis en relief des traits importants de la faiblesse de volonté par des expérimentations animales. Il est difficile de dire cependant si ce manque d'intérêt découlerait d'une sorte de spectre moral qui planerait encore sur les conceptions philosophiques de la motivation humaine.

forme d'ignorance et une erreur de mesure. Ce point est particulièrement intéressant parce qu'il s'accorde, dans une certaine mesure et sous certaines conditions, avec la théorie contemporaine du choix intertemporel et de l'escompte hyperbolique du futur, que je présenterai dans le chapitre 9.

Aristote n'a pas dégagé comme Platon le problème de la faiblesse de volonté du domaine moral. Pour le Stagirite, les lignes de conduite, dans lesquelles s'engagent les agents à la volonté faible, sont sujettes au blâme (*Éthique à Nicomaque* : 1110a19). Par contre, bien qu'il ait défendu, à l'encontre de la position de son maître, l'existence de la stricte akrasie, Aristote offre une théorie beaucoup plus complexe de la motivation et, à cet égard, un peu plus touffue que celle de Platon. Mais il y a des passages assez clairs sur les processus de délibération et les causes putatives de la faiblesse de volonté en général – et pas seulement sur la stricte akrasie. Aristote identifie l'akrasie avec le choix impulsif qu'il distingue du choix réfléchi et classe les choix akratiques dans la catégorie du premier. Cela s'accorde d'ailleurs avec nos intuitions selon lesquelles les agents irréfléchis, comme les enfants en bas âge, par exemple, sont plus susceptibles de vivre des épisodes de faiblesse de volonté.

Par ailleurs, Aristote offre une explication de la faiblesse de volonté qui a fait boule de neige et dont la proposition de base est aussi très intuitive. *Grosso modo*, Aristote voit dans l'émotion non enrégimentée ou non réformée par la raison la cause de la faiblesse de volonté. Aussi, les états émotifs sont typiquement vus comme un facteur potentiel d'irrationalité pratique. Si l'homme écoute son cœur et non sa raison, il est davantage susceptible de se montrer faible. Même si Aristote élabore un appareil théorique complexe et sophistiqué pour expliquer la possibilité de la stricte akrasie et la faiblesse de volonté en général, le résultat s'accorde assez bien avec certaines de nos intuitions communes et préanalytiques sur les causes du phénomène, à savoir qu'en chacun de nous vit une sorte de pulsion irrationnelle à agir à l'encontre de nos intérêts et des principes moraux qu'on accepte pourtant, et qu'on doit harnacher par la force de notre raison.

Je vais présenter et examiner maintenant certaines propositions que Platon et Aristote ont faites à l'égard de la faiblesse de volonté et de la stricte akrasie. Je vais mentionner ensuite certaines critiques qu'on peut leur adresser.

2.1.1 La conception platonicienne de la motivation

L'identification du problème de la faiblesse de volonté et ses premières explications théoriques par la tradition gréco-latine furent d'emblée campées dans le domaine moral. *Video meliora proboque – deteriora sequor* : « Je vois le bien, je l'approuve, et je fais le mal » affirmait la Médée d'Ovide (*Métamorphoses* : VII, 20). De même, l'*Hippolyte* d'Euripide comprend un passage qui identifie le même problème :

> « *Le bien, nous le connaissons, notre raison le conçoit. Mais nous nous rebutons devant l'effort qu'il exige, soit qu'il contrarie notre mollesse, soit qu'aux plaisirs de la vertu nous en préférions quelque autre.* ». (*Hippolyte porte-couronne* : 380-384)

La première ébauche d'explication de ce phénomène se trouve dans la doctrine du cognitivisme moral de Platon. Dans le *Ménon*, Socrate soutien que

> « *les gens [méchants] ne désirent pas le mal, puisqu'ils ignorent ce qu'il est, mais qu'ils désirent ce qu'ils croyaient être le bien, même si en fait ce bien est mal. De sorte que, s'ils ignorent le mal et le prennent vraiment pour un bien, il est évident que c'est le bien qu'ils désirent.* » (*Ménon* : 78a)

Ainsi, pour Platon les agents moraux ne peuvent faire le mal volontairement, c'est-à-dire qu'ils ne peuvent accomplir des actes moralement répugnants que s'ils ignorent des faits pertinents ou qu'ils y sont contraints par une force extérieure. L'hypothèse d'un bagage informationnel inadéquat, comme facteur exclusif de faute morale dans les contextes où l'agent est libre, a fait couler beaucoup d'encre et figure comme la première clef pour comprendre le phénomène de la faiblesse de volonté. En effet, Platon calque son approche du problème de la motivation en général sur celui du problème moral. Dans *Le Protagoras*, Socrate identifie la poursuite du Bien avec la recherche du plaisir. Il rejette l'idée que les agents qui vivent un épisode de faiblesse de volonté sont en fait « vaincus » par les plaisirs. Au contraire, pour Socrate il faut rechercher à maximiser son plaisir dans les limites de ce que commande la justice :

> « […] *Si tu pèses l'agréable avec l'agréable, il faut toujours choisir le plus grand et le plus nombreux ; si tu pèses le pénible avec le pénible, il faut choisir le moindre et le plus petit ; si tu pèses l'agréable avec le pénible, et si l'agréable est en excès par rapport au désagréable, que le long terme l'emporte sur l'immédiat ou l'immédiat sur le long terme, tu dois agir en respectant toutes ces données ; si le désagréable est en excès par rapport à l'agréable, il ne faut rien faire.* » (*Protagoras* : 357-358d)

Le problème de la faiblesse de volonté réside dans le fait que les agents échouent à identifier les lignes de conduite qui maximisent effectivement notre plaisir global. Par une série d'analogies éclairantes, Socrate assimile cette difficulté à des phénomènes

d'illusions et dont la solution réside dans la pondération informée des options qui s'offrent à nous :

> « […] *S'il en est ainsi, répondez-moi, dirai-je. Les mêmes grandeurs vous paraissent, à la vue, plus grandes de près, et plus petites de loin, ou non? – ils diront que oui. – Même chose pour les épaisseurs et quantités? Et même des sons équivalents paraissent plus forts de près, et moins forts de loin? – Ils en conviendraient. – Si donc nous considérions que l'activité réussie consistait à choisir dans nos actions les grandes dimensions, et d'éviter dans nos actions les petites dimensions, qu'est ce qui, manifestement, assurerait la sauvegarde de notre vie? Est-ce l'art de la mesure ou la puissance de l'apparence? Ne faut-il pas dire que celle-ci nous égarerait, nous ferait souvent tout mettre sens dessus dessous, nous conduirait à nous repentir, dans nos actions et dans nos choix du grand et du petit, alors que l'art de la mesure rendrait cette illusion sans force, et, en faisant connaître le vrai, prodiguerait le repos de l'âme qui trouverait sa stabilité dans le vrai et sauverait sa vie? Les hommes accorderaient-ils que c'est l'art de la mesure qui nous sauve en cela, ou que c'en est un autre?* » (*Protagoras* : 357-358d)

Beaucoup de commentateurs ont soutenu que Platon nie en fait l'existence de la faiblesse de volonté. Or, c'est une position exégétique controversée. Tout dépend de la conception qu'on a de la faiblesse de volonté et de la manière dont on peut s'en servir pour interpréter les propos de Socrate. Si notre conception de référence est que la faiblesse de volonté est une sorte de changement d'avis irrationnel à l'égard des options qui s'offrent à nous (faiblesse de volonté diachronique ou « aveugle », akrasie tout court), alors on peut dire que Platon accepte la réalité du phénomène. En revanche, si notre conception de référence est la stricte akrasie, alors on a plus de raisons de penser que Platon en nie la possibilité. Comme je l'ai mentionné dans l'introduction, la stricte akrasie consiste pour un agent à agir à l'encontre de son meilleur jugement le plus récent. Or, la conception classique – ou la plus répandue – de la faiblesse de la volonté est justement celle de la stricte akrasie, ce qui fait croire que Platon en niait l'existence. Certains commentateurs (Vlastos, 1956), après avoir distingué les deux conceptions de l'akrasie, prétendent que Platon défendait une conception diachronique et niait la stricte akrasie. D'ailleurs, ces mêmes commentateurs attribuent cette interprétation au Stagirite quand celui-ci critiqua la position de son maître.

Les raisons qui motivent Platon à défendre une forme de cognitivisme motivationnel dans l'explication de la faiblesse de volonté reposent sur la distinction entre *opinion* et *connaissance*. En parlant des opinions vraies, Socrate soutient que :

> « […] *aussi longtemps qu'elles demeurent en place, sont une belle chose et tous les ouvrages qu'elles produisent sont bons. Mais ces opinions ne consentent pas à rester longtemps en place, plutôt cherchent-elles à s'enfuir de l'âme humaine; elles ne valent donc pas grand-chose, tant qu'on ne les a pas reliées par un raisonnement qui en donne l'explication. Mais*

> *dès que les opinions ont été reliées, d'abord elles deviennent connaissances, et ensuite, elles restent à leur place.* » (*Ménon : 98a)*

Une fois qu'elle se trouve implémentée dans l'esprit, la connaissance rend les agents résistant aux épisodes de faiblesse de volonté (*Protagoras* : 352d7 - E2, 352 E6, 353a5-C2, 357c7-d1, E2). L'instabilité de l'opinion – ou croyance non rationnellement justifiée – explique pourquoi les agents sont inconstants et changent d'avis aisément sur les lignes de conduite qu'il convient d'adopter. Comme la connaissance est censée toujours identifier le bien le plus grand, il s'ensuit que les agents qui vivent ces épisodes sont mus par l'opinion et non par la connaissance.

2.1.1.1 La perspective temporelle et le plaisir d'anticipation dans le Philèbe

Dans le dialogue *Le Philèbe*, Platon fournit des détails supplémentaires pour comprendre son intellectualisme motivationnel. Il introduit la notion de *plaisir faux*[4] pour expliquer les erreurs de jugement qui poussent à l'occasion les agents à adopter des comportements relevant de la faiblesse de volonté. Les plaisirs faux ne sont rien d'autre que des plaisirs que l'agent ressent présentement à l'idée fausse de ressentir des plaisirs futurs. Dans ce dialogue, Platon distingue donc le plaisir d'anticipation du plaisir anticipé (*Le Philèbe* : 39d-41a). Ce faisant, il raffine sa position sur ce qui constitue selon lui le « *bottom-rock* » de la motivation, à savoir le plaisir. Si les agents rationnels cherchent à maximiser leur plaisir (dans les limites de la justice), on ne doit pas seulement tenir compte du plaisir qui découlera des conséquences de leurs choix *dans le futur*, mais également du plaisir qu'ils éprouvent dans le présent à l'idée que ces conséquences se réaliseront. Or, il arrive que le plaisir d'anticipation ne soit pas proportionnel au plaisir anticipé. Dans ce cas, l'agent échouera à maximiser son profil hédonique. Il pourrait, par exemple, sous-estimer la quantité de plaisir, qu'il éprouvera dans un futur lointain, découlant du fait d'investir maintenant des capitaux dans un fond de retraite, *parce qu'il ne ressent pas présentement suffisamment de plaisir à l'idée de disposer d'un bon fond de retraite*. De même, il pourra également surestimer le plaisir, qu'il éprouvera dans le futur, découlant de l'achat actuel d'une voiture de luxe, qu'il a à peine les moyens de se payer, parce qu'il ressent trop de

[4] À ne pas confondre avec la notion de *faux plaisirs* désignant des épisodes affectifs positifs illusoires comme le sont les manifestations affectées de plaisir dans un contexte où aucune cause réelle du plaisir n'est réalisée.

plaisir présentement à l'idée de disposer d'une voiture de luxe dans le futur en dépit du fait qu'il devra se serrer la ceinture.

Bien qu'il ne soit pas clairement exprimé dans le texte du dialogue, Ronald de Sousa suggère d'appeler *Principe de Philèbe* la maxime pratique qui veut que lorsque nous nous apprêtons à faire un choix dont les considérations hédoniques priment, nous devons nous assurer que « *le plaisir d'anticipation soit toujours proportionnel au plaisir anticipé* » (De Sousa, 2000). De Sousa fait des plaisirs d'anticipation des sortes de méta-émotions qui portent sur d'autres plaisirs. Cependant, il semble que Platon ne va pas jusque-là et se contenterait plutôt d'affirmer que les faux plaisirs sont des opinions que l'agent akratique entretient présentement sur des plaisirs escomptés. D'ailleurs, Platon n'examine pas dans ce dialogue l'hypothèse d'une confusion de l'agent akratique, qui prendrait ses plaisirs anticipant pour des plaisirs anticipés au moment où il choisit une ligne de conduite désavantageuse. La cause semble résider exclusivement dans l'estimation erronée de l'intensité et de la qualité des états hédoniques futurs. Ce qui reste tout à fait conforme à son intellectualisme motivationnel.

2.1.1.2 Que penser du cognitivisme motivationnel?

Il est naturel de penser que le cognitivisme motivationnel de Platon est pour l'essentiel calqué sur sa conception du jugement moral. Les agents font des choix irrationnels pour les mêmes raisons qu'ils errent au niveau moral. Il n'a donc qu'en partie dégager la faiblesse de volonté de la perspective morale. Dans tous les cas, les agents ignorent certaines informations pertinentes, font de mauvaises inférences ou utilisent des outils de mesure inappropriés.

Or, le cognitivisme (réalisme ou rationalisme) moral est apparu à plusieurs comme une doctrine naïve – du moins dans sa version platonicienne. En fait, les arguments qui vont à l'encontre de celui-ci s'appliquent également au cognitivisme motivationnel. Il ne suffit pas d'éclairer coquins et méchants pour que ceux-ci deviennent *ipso facto* disposés à réformer leur conduite, même dans les situations où ils *comprennent* le sens et la portée de ce qu'on leur dit. De même, il ne suffit pas d'éclairer les personnes à propos des conséquences néfastes de telle ou telle décision ou habitude pour qu'ils réforment leur conduite de sitôt.

Je ne compte pas ici relater l'état des débats actuels sur la plausibilité du cognitivisme moral[5]. Aussi, je n'en dirai que ce qui est nécessaire pour comprendre, je pense, la position de Platon à l'égard de la motivation en général.

Le cognitivisme moral semble intuitivement erroné notamment parce que les débats moraux s'enlisent souvent dans des conflits très polarisés où les parties prenantes ne convergent pas vers une solution au fur et à mesure que leur position se définie et se raffine. Ces débats semblent s'enraciner dans un conflit de valeurs si profond qu'aucune solution n'apparait envisageable. Cependant, quiconque discute d'un problème moral avec quelqu'un qui tient une position différente de la sienne s'aperçoit, s'il a suffisamment de recul analytique, que ce qui empêche tout le monde de converger vers une solution réside la plupart du temps dans un désaccord sur les faits et rarement dans un désaccord sur des valeurs (Putnam, 2004). Même dans les cas où les désaccords reposent réellement sur une divergence de valeurs, il existe tout de même un espace de délibération pour questionner l'exigence d'inclure certaines valeurs au nombre des raisons pertinentes pour résoudre un problème moral ou pour discuter de la place que ces valeurs doivent occuper dans une échelle de valeurs pertinentes, ou pour déterminer le poids relatif des valeurs pour trancher en cas de conflit de valeurs. Dans chacun de ces niveaux de discussion, on fait intervenir des considérations factuelles nombreuses. Il y a des antécédents cognitifs multiples et variés à l'adhésion à des valeurs comme la liberté, la sécurité, l'excellence, le courage, la quantité plus grande d'opportunités, l'intelligence, la santé, la reconnaissance sociale, etc. De même qu'il y a, pour les agents des antécédents cognitifs multiples et variés à l'adhésion à des échelles de valeurs, qu'elles soient contextuelles ou absolues.

Accepter cela n'implique cependant pas que tous les conflits moraux se règleront demain matin si tout le monde accepte de bonne foi qu'ils reposent sur une ignorance de certains faits pertinents. Les débats scientifiques perdurent souvent pendant de nombreuses années bien qu'ils ne reposent pas (ou peu) sur des divergences de valeurs[6]. D'ailleurs, toutes sortes de superstitions, de préjugés et de croyances fausses sont plus coriaces que beaucoup de positions morales, en dépit du fait qu'ils reposent sur des considérations factuelles

[5] Voir l'intéressant recueil *Le réalisme moral* (Ogien, 1999).

[6] Bien que les valeurs jouent également un rôle dans la discussion, la sélection et le rejet des hypothèses et théories scientifiques. Mais il s'agit évidemment de valeurs épistémiques et non morales, comme la simplicité, la force prédictive et explicative, l'élégance, l'opérationnalité, la conformité avec des hypothèses existantes, etc.

inadéquates. La discussion de telles croyances tourne souvent à vide parce qu'il est souvent difficile d'identifier les véritables raisons qui motivent l'adhésion de certains agents à celles-ci. Aussi, il est difficile de leur couper l'herbe sous le pied. Mais quand nous arrivons à les identifier et à montrer à l'agent que ces raisons ne sont pas acceptables ou pas suffisantes pour en inférer sa croyance, cela augmente la probabilité qu'il révise son jugement. Et l'on n'a pas de raison de croire qu'il n'en va pas de même pour les jugements moraux.

Cela dit, le cognitivisme moral se heurte cependant à une difficulté importante. Il semble qu'il y ait des choses qu'on doive valoriser pour elles-mêmes et qui doivent faire l'objet d'aucun compromis avec d'autres valeurs : le bien-être, le bonheur, la joie, la satisfaction, la plénitude ou le contentement. Notre sensibilité à l'égard de la condition affective d'autrui est une condition minimale pour avoir des attitudes morales. C'est pourquoi l'analyse des arguments évoqués dans une discussion morale tend à montrer que les agents moraux partagent une sorte de « *bottom-rock* » normatif indiscutable, comme l'idée qu'on doive valoriser le bien-être individuel et collectif lorsque vient le temps de promouvoir ou de défendre des institutions et des modes de vie. Ces normes de base n'ont pas d'antécédents cognitifs, mais sont enracinées dans une disposition affective de nature altruiste. Bien que cela reste encore débattu parmi les spécialistes de biologie évolutionniste, il semble que même dans sa forme minimale – qui consiste en une non-indifférence à l'égard de la souffrance d'autrui – une disposition altruiste est une condition nécessaire aux comportements moraux authentiques (*full-fledged*). Les agents qui n'ont pas cette sensibilité minimale restent imperméables aux arguments moraux. Si ces agents éprouvent en plus un irrépressible plaisir à dominer complètement autrui, ils risquent d'adopter des comportements psychopathiques[7].

[7] Bien que les données sur le sujet fassent encore l'objet de polémiques, beaucoup de chercheurs en neurosciences expliquent cette désensibilisation par un dysfonctionnement des neurones miroirs spécialisés dans la simulation des états émotionnels d'autrui (Frans de Waal, 2009 ; G Rizzolatti, L Folgassi & V Gallese, 2007). Si l'on est indifférent à la situation affective d'autrui et qu'on éprouve même du plaisir à dominer physiquement et intégralement, on ne cherchera pas à faire le mal pour le mal, mais on s'engagera dans une ligne de conduite aussi moralement répugnante soit-elle – comme des actes de torture et de séquestration. Certains psychopathes savent néanmoins que leurs comportements, en dépit du plaisir qu'ils en retirent, sont moralement répréhensibles. Seulement, le déplaisir produit par les neurones miroirs et par la honte subséquente n'est pas suffisamment important pour que cela motive ces agents à s'abstenir d'accomplir de tels actes. Il y a pour ainsi dire un compromis entre, d'une part, le déplaisir que procure un acte entendu comme moralement répugnant et, d'autre part, le plaisir qui découle de certains de ses aspects, mais nettement en faveur de ces derniers, en dépit de l'ampleur du mal qui est produit.

L'absence d'une sensibilité minimale à l'égard de la condition affective d'autrui peut cependant être envisagée comme l'absence d'une source d'information essentielle pour établir une connaissance robuste sur ces états affectifs. Aussi, le comportement des psychopathes serait dû en partie à un manque de connaissances.

Or, cette hypothèse semble démentie par l'étude du cerveau des psychopathes. Une étude qui porte sur la capacité des psychopathes à inférer des états émotionnels à partir de situations où les agents réussissent à réaliser ou non leurs intentions, tend à montrer que ces premiers sont capables de raisonner sur les états émotionnels lorsqu'une évaluation explicite de ceux-ci leur est demandée. Bien que les psychopathes cobayes répondent de manière inadéquate aux états émotionnels des autres personnes, ils sont néanmoins capables d'identifier ceux-ci aussi bien que les personnes normales. Ce n'est donc pas un déficit dans la connaissance de la condition affective d'autrui qui explique pourquoi les psychopathes y réagissent de manière si peu adéquate. En observant avec un appareil de résonnance magnétique (IRMf) les circuits de neurones qui contribuent à cette identification, on se rend compte que ceux-ci ne sont pas les mêmes que ceux que les personnes normales activent. Chez ces dernières, outre le système des neurones-miroirs, deux régions, soit le gyrus supramarginal dans le cortex supérolatéral et le gyrus supérieur dans le cortex préfrontal (région impliquée dans le phénomène de conscience de soi), sont activés pour accomplir cette tâche. Or, dans le cerveau des psychopathes les chercheurs ont observé que ce sont plutôt des régions associées à l'attention et au monitorage des actions en fonction d'un résultat, soit le cortex orbito-frontal, le cortex frontal médian et les aires temporo-pariétales, qui sont sollicitées pour identifier les émotions d'autrui. Comme le disent Sommer et coll., les auteurs de l'étude,

> « […] *bien que les psychopathes ne montrent pas de déficits lorsqu'ils raisonnent à propos des émotions des autres personnes, quand on leur demande explicitement de le faire, ils utilisent des stratégies de processus neuronaux qui sont plus reliés à la rationalité et aux processus conséquentialistes (*outcome-oriented processes*).* » (2010 : 92)

Comme les psychopathes n'ont, en général, pas de carences cognitives – en fait, beaucoup d'entre eux réussissent mieux à des tests d'intelligence variés que les personnes normales – biaisant systématiquement leurs jugements à la faveur de croyances irrationnelles, et ce, même dans des domaines spécifiques, leur cas discrédite en partie le cognitivisme moral.

Bien que le sujet soit controversé et qu'il y ait toute une littérature qui en traite, je ne compte pas en discuter ici. Mais on peut penser que le cognitivisme moral est erroné parce qu'il y a un fondement émotionnel qui ne repose pas sur des considérations factuelles plus fondamentales aux yeux des agents[8]. On remet en question les choix collectifs en mettant en relief le fait qu'ils ne sont pas de bons moyens pour atteindre le bien-être d'au moins une partie de la population ou qu'il existe des moyens plus appropriés. On ne remet pas en question cependant l'objectif d'atteinte du bien-être social. Cela n'implique cependant pas que les sociétés gardent *de facto* toujours le cap sur cet objectif. Celui-ci peut être oublié en cours de route, souvent quand on met l'emphase sur les moyens institutionnels plutôt que les fins ou les indicateurs de bien-être (ex. : le tout à l'économie et à l'accroissement du PIB national).

Bien qu'il soit communément admis que Platon défendait une forme de cognitivisme moral, il est plus difficile de savoir avec certitude s'il loge à la même enseigne pour ce qui est de la motivation en général. Je pense qu'il est assez naturel de penser que c'est le cas, et que sa conception du jugement pratique est une extension de la conception du jugement moral. C'est cependant une question technique d'exégèse que je veux laisser ouverte aux fins de cet exercice.

Mais que doit-on penser du cognitivisme motivationnel apparemment adopté par Platon? Il y a tout lieu de croire que les difficultés que rencontre le cognitivisme moral sont aussi rencontrées par le cognitivisme motivationnel. Les considérations qui motivent en fin de compte les agents ne relèvent pas à proprement parler de leurs croyances ou bagages cognitifs. Il n'y a pas d'antécédents cognitifs pour motiver de manière générale la poursuite du bonheur, d'état de satisfaction, de félicité, de paix intérieur, ou toute autre forme de bien-être. Un énoncé du genre : « Ça serait bien pour moi d'être heureux » sonne plus comme une tautologie que comme un énoncé informatif. Aussi, le plaisir, qu'il soit dans ses formes les plus sophistiquées (plaisir d'accomplissement, de reconnaissance, satisfaction de justice rendue, etc.) ou les plus élémentaires (plaisir d'uriner, d'absorber des matières comestibles, de réaliser un état homéostatique optimal, etc.), constitue un objectif dont les agents estiment avec raison ne pas avoir à justifier l'attrait. La seule justification rationnelle

[8] Même s'il est judicieux de faire comme s'il était vrai dans les débats à saveur morale.

concernant la poursuite des plaisirs porte davantage sur la façon de les réaliser et sur la manière de les distribuer dans les temps, que ce pourquoi on doit les réaliser.

Socrate nous exhorte certes à rechercher le maximum de plaisir – dans les limites de la justice – sans fournir de raison plus fondamentales. Cela laisse croire que des états émotionnels forment le soubassement de la motivation et non des états cognitifs. Mais il suggère en même temps que les agents autonomes et maîtres d'eux-mêmes peuvent prendre suffisamment de recul à l'égard des constituants affectifs et désidératifs, que sont le plaisir et l'attrait du plaisir pour choisir parmi eux. Aussi, le rôle de la raison consisterait, dans cette optique à assigner de manière la plus correcte possible – conformément à l'idéal d'objectivité – des poids à ces constituants. C'est pourquoi ces derniers ne forment pas vraiment le *bottom-rock* motivationnel des agents. C'est plutôt en amont, dans la *formation du jugement* à l'égard de l'intensité et de la quantité des plaisirs, que se situe, pour Platon, ce « *bottom-rock* » motivationnel. Les considérations affectives et désidératives jouent bien un rôle dans l'explication des motivations d'un agent, mais seulement sous forme d'exhortations (*hortatory urge*). C'est toujours la raison qui a le dernier mot, qui même choisit de prêter attention ou non aux « demandes » des systèmes affectifs et désidératifs. En simplifiant un peu les choses, le psychologue Ainslie soutient que « […] *cognitives theories say that emotion/motivation just provide one more challenge for reason to meet. If we succumb to temptation, the problem lies with our reasoning.* » (2001 : 15)

Pour Ainslie, le problème avec ce genre de conception est que le processus consistant à donner du poids à ses désirs reste mystérieusement indéterminé. Si donner du poids à des désirs n'est pas lui-même un processus déterminé en amont par des désirs, alors cela met fin, d'une certaine manière, à l'analyse du choix. S'il n'y a pas de manière d'établir un lien entre les facteurs affectifs et désidératifs, d'une part, et le jugement, d'autre part, alors on devra accepter l'idée qu'il y a une sorte de petit homoncule dans l'esprit qui tantôt incline du côté des désirs et émotions, tantôt non, mais dont on ne peut pas expliquer les choix. Or, non seulement il s'agit d'une position théorique stérile, mais on sait maintenant que même les processus cognitifs les plus froids – intervenant dans la délibération pratique – sont orientés par des mécanismes affectifs et désidératifs plus fondamentaux. Par exemple, la formation du jugement dans un contexte de choix est en partie déterminée par des mécanismes d'attention, lesquels sont en partie déterminés par des considérations affectives

et désidératives. Le simple fait que notre esprit enclenche un processus délibératif pour résoudre un problème comme, par exemple, le fait de savoir si je dois commettre l'adultère avec une ou plusieurs autres femmes est déterminé en partie par l'attention involontaire que je porte aux corps féminins que je croise dans mes déambulations. Si je ne portais jamais attention aux femmes, ou à certains de leurs traits, je n'entamerais jamais de processus décisionnels ayant l'adultère comme enjeu. Aussi, si j'étais insensible aux charmes féminins, si j'étais un homosexuel « intégriste », je n'y porterais même pas attention. Le processus décisionnel n'est pas seulement enclenché par des considérations affectives et désidératives, mais est noyauté du début à la fin par ces considérations.

Cela dit, les critiques qu'on adresse spontanément au cognitivisme motivationnel, comme celles qu'on adresse spontanément au cognitivisme moral, ne sont pas correctes pour autant. Elles reposent souvent sur des considérations intuitives et restent à la surface du phénomène du processus décisionnel. On dira, par exemple, qu'on peut savoir que le cancer guette si l'on n'arrête pas de fumer et continuer en dépit de cette croyance ou encore estimer qu'on est en pleine période de procrastination alors qu'on sait devoir rédiger un travail important. Aussi, l'acquisition d'informations pertinentes ne semble pas toujours suffisante pour qu'on décide d'emprunter une ligne de conduite plus adéquate. Un individu peut bien continuer à fumer en dépit du fait qu'il connaisse parfaitement bien les conséquences désastreuses du tabagisme. Un autre peut s'adonner à la procrastination tout en connaissant les impacts négatifs que cela aura sur son bien-être futur. Dans les deux cas, le savoir ne semble pas une donnée manquante. Les individus possèdent toutes les connaissances nécessaires pour l'adoption de la meilleure ligne de conduite, et cela apparaît indubitable.

C'est ce qu'on aimerait croire, et pourtant, plusieurs données expérimentales tendent à montrer que les agents ont une conception relativement peu claire de la qualité, de l'intensité et de la durée de leurs états hédoniques futurs. Dans beaucoup de contextes de choix, les agents éprouvent beaucoup de difficultés à faire des prédictions fiables sur leurs états affectifs futurs. Aussi, les décisions qu'ils prennent en fonction des conséquences futures reflètent souvent une mauvaise anticipation de leurs propres états mentaux corrélatifs. Kassam, Gilbert, Boston et Wilson ont montré (2008), par exemple, comment des participants appartenant à des groupes mixtes d'hommes et de femmes anticipent leur

niveau de bien-être futur dans un contexte où ils doivent choisir entre un gain monétaire disponible maintenant et un gain disponible dans 3 mois. En ayant exclu l'hypothèse que les agents confondent dans leur mesure subjective leurs états affectifs futurs avec leurs états présents au moment de choisir une option, et en ayant exclu aussi l'hypothèse que les agents estiment que la seule prédiction de leurs états futurs modifiera en retour ces derniers, les résultats attestent l'hypothèse que *les agents évaluent à la baisse l'étendue et l'intensité de leurs états hédoniques futurs*. Cela va à l'encontre de l'hypothèse standard de l'attitude des agents vis-à-vis le futur, à savoir « […] *that there is something wrong with people's* decisions *about the future but nothing wrong with their* perception *of it*. »[9] (2008 : 1537)

Je ne discuterai pas ici du problème plus général de la prédiction de nos propres états affectifs futurs (*affective forcasting problem's*). Je compte discuter de manière plus détaillée de ce problème au chapitre 9. Ce qui est, je pense, important de retenir ici est qu'on ne doit pas rejeter la conception platonicienne de la motivation sur la base d'intuitions fortes concernant la capacité des agents à mesurer correctement l'intensité et la quantité de bien-être qui découle des choix qui doivent être faits. D'ailleurs, la difficulté des agents à prédire leurs propres états affectifs futurs peut nous servir à interpréter le principe de Philèbe, du moins à expliquer pourquoi les agents akratiques ne semblent pas s'y conformer. Si les agents akratiques ne respectent pas ce principe, c'est-à-dire s'ils ne prennent pas soin de moduler leur état affectif présent de manière à ce qu'il soit proportionnel à leurs états affectifs futurs avant de faire un choix, c'est peut-être parce qu'ils mesurent incorrectement ces derniers (états futurs). Aussi, les antécédents cognitifs d'un agent akratique pourraient jouer un rôle plus important dans l'explication de ses choix que nos intuitions nous portent à le croire.

Cela dit, il ne s'ensuit pas que le cognitivisme motivationnel soit correct. Les agents peuvent avoir des difficultés importantes à estimer leur bien-être futur et que cela explique en partie le fait qu'ils soient disposés à prendre des décisions désavantageuses, mais que la recherche de ce bien-être est motivée par des considérations désidératives et affectives plutôt que cognitives.

[9] *Ibid*. p. 1537. Souligné par les auteurs.

En fait, le tableau est plus complexe qu'il n'y parait. On ne peut pas non plus défendre l'idée que la motivation relève strictement des émotions et/ou des désirs. On peut défendre en même temps l'idée que les paramètres motivationnels les plus fondamentaux relèvent d'états hédoniques de base ayant une saveur émotionnelle particulière et/ou des désirs visant leur réalisation, et accepter en même temps l'idée que les désirs et les émotions de niveau supérieur ont des antécédents cognitifs. C'est d'ailleurs en identifiant ces antécédents qu'on arrive à critiquer les désirs et les réactions émotionnelles d'autrui ou les nôtres. On dira par exemple à un ami qu'il ne devrait pas désirer acheter telle ou telle maison parce qu'il estime à tort qu'elle est dans un quartier tranquille ou quelle est en bon état ou qu'il n'y a pas de meilleure candidate sur le marché. Mais on ne critiquera pas son désir d'améliorer son sort ou d'être plus heureux en achetant une maison. De manière analogue, les émotions complexes comme la colère ont des antécédents cognitifs qu'on peut mettre en relief lors d'un examen critique. On peut juger que la colère d'une personne à l'égard d'une autre est inappropriée parce qu'elle repose sur une interprétation erronée de ses intentions ou qu'elle est disproportionnée étant donné la gravité du geste. Mais on ne critiquera pas le plaisir qu'elle éprouve à uriner ou à ingurgiter des aliments sucrés – du moins, pas en identifiant des antécédents cognitifs.

Pour peu qu'ils soient complexes ou situés à des niveaux supérieurs, les émotions et les désirs ont des antécédents cognitifs, et c'est sans doute ce qui nous inciterait à voir dans le cognitivisme motivationnel une conception plausible. Toutefois, les facteurs cognitifs constituent un paramètre de choix parmi d'autres, à côté des paramètres désidératifs et affectifs. On peut modifier le comportement d'un agent en jouant sur l'un ou l'autre de ces paramètres. Mais on peut penser qu'il y a des paramètres affectifs et désidératifs de base suffisamment fixes et essentiels au maintien des fonctions vitales qui constituent le soubassement de la motivation des agents pour que certaines modifications soient difficiles, voire impossibles à réaliser[10].

[10] Par contre, on peut accepter l'idée que des facteurs cognitifs modulent à l'occasion des états émotionnels aussi simples que celui d'avoir envie d'uriner. Elster a mis en relief le fait qu'apprendre qu'il y a des lieux appropriés pour uriner dans l'environnement immédiat d'une personne peut accentuer grandement son envie d'uriner (Elster, 2007). Mais, dans ce cas, ce n'est pas parce que la personne apprend qu'il y a un cabinet de toilette dans les parages qu'elle ressent l'envie d'uriner, mais c'est pour cette raison qu'elle en ressent davantage l'envie.

Je décrirai plus en détail dans les chapitres 3 et 4 la nature des motivations et leurs fonctions paramétriques. J'aborderai au chapitre 7 la question de l'introduction d'un paramètre supplémentaire (règles procédurales) ignoré de la tradition philosophique, mais qui nous permet d'envisager des explications différentes de la faiblesse de volonté et de formuler une critique importante de la conception standard du phénomène.

2.1.1.3 Une conception normative?

Le rôle que joue le principe de Philèbe dans les exposés de Socrate semble impliquer que la conception platonicienne est une conception normative et non pas seulement descriptive. Dans une conception normative, un diagnostic de faiblesse de volonté implique nécessairement que l'agent a échoué à se conformer à une norme pratique ou épistémique essentielle – et non pas de manière accessoire. Parmi les promoteurs actuels d'une conception normative, on retrouve Bratman et Holton, dont je présenterai les conceptions au chapitre 8. Mais aussi dans une certaine mesure Donald Davidson qui a proposé de faire du principe de continence (« agis toujours de manière à être conforme à ton meilleur jugement ») un principe constitutif de la rationalité pratique que l'agent akratique enfreint nécessairement (Davidson : 1970).

La conception de Platon peut être vue comme étant normative parce que, pour lui, l'agent akratique échoue à se conformer à une norme pratique : le principe de Philèbe. Ce principe, comme je l'ai mentionné, nous enjoint à nous assurer que le plaisir d'anticipation doit être proportionnel au plaisir anticipé quand on évalue des options dont les conséquences sont distribuées dans le temps. Or, l'agent akratique échoue d'une manière particulière à appliquer ce principe : il choisit alors que son plaisir d'anticipation est soit trop grand soit trop petit. Platon ne nous donne cependant pas beaucoup de détails ici. Il est permis de penser que Platon estimait que le plaisir d'anticipation a tendance à être trop grand lorsque l'agent considère des activités plaisantes, mais dont le résultat est moins plaisant (ex. : la surconsommation alimentaire), et trop petit lorsqu'il considère des activités déplaisantes, mais qui conduit à un résultat plaisant (ex. : les besognes). Platon ne nous dit pas cependant si nous devons considérer que tous les épisodes de faiblesse de volonté consistent nécessairement dans une forme de disproportion des deux types de plaisirs, laquelle est proscrite par le principe de Philèbe. Si tel est le cas, alors Platon propose une conception normative de la faiblesse de volonté. Mais il est évident que ce principe est un principe

pratique qui s'inscrit dans l'idéal platonicien de recherche de l'autonomie : appliquer ce principe est une manière de brider ses passions dans la poursuite du Bien.

Cela dit, le fait que Platon défende une forme de cognitivisme motivationnel nous amène à penser qu'il n'y a pas pour lui que des normes pratiques qu'on enfreint nécessairement lorsqu'on choisit une ligne de conduite akratique, mais aussi des normes épistémiques. L'agent akratique ne fait pas que s'éloigner de l'idéal d'autonomie, il s'éloigne également de l'idéal d'objectivité. Il s'éloigne de ce dernier parce qu'il ne se guide que par l'opinion et non par la connaissance. Beaucoup de choses ont été dites sur la distinction entre opinion et connaissance[11] et le fait de la faire intervenir peut facilement obscurcir le propos (Davidson : 1970). Mais ce qui est pertinent de mentionner ici est qu'un agent a des devoirs épistémiques conformes à l'idéal d'objectivité, entre autres celle de s'efforcer d'avoir des croyances vraies *et* rationnellement justifiées. Or, il s'avère que l'agent akratique, en ne se guidant qu'avec ses opinions, ne s'acquitte pas de son devoir épistémique minimal. Aussi, l'agent akratique agit en fonction de raisons, mais pas de bonnes raisons.

Être motivé par des opinions apparait chez Platon comme une condition nécessaire, toutefois pas suffisante pour l'identification d'un épisode de faiblesse de volonté. Les opinions au sens platonicien motivent des tas de comportements qui n'ont rien à voir avec la faiblesse de volonté. Le non-respect du principe de Philèbe est, à cet égard, un trait constitutif plus adéquat pour servir de critère normatif[12].

---

[11] Dans la littérature exégétique, mais également en épistémologie, notamment depuis la publication du célèbre article de Gettier (1963) qui mettait en relief un problème conceptuel concernant une version moderne de la distinction, celle entre croyance vraie et croyance vraie rationnellement justifiée.

[12] L'idée platonicienne d'intégrer des considérations normatives épistémiques dans la conception de base de la faiblesse de volonté a, contrairement aux considérations normatives pratiques, quelque chose de peu intuitif si l'on n'accepte pas le cognitivisme motivationnel. Il est difficile de soutenir que l'agent qui vit un épisode de faiblesse de volonté n'a pas, quelque part, fait ses devoirs épistémiques, et cela, en dépit du fait qu'il puisse à l'occasion mal évaluer la qualité, l'intensité et la quantité de ses états hédoniques futurs. En revanche, il ne fait pas de doute que *beaucoup* d'épisodes de faiblesse de volonté sont motivés par des *croyances motivées* que l'agent *ne devrait pas avoir*, et cela donne un certain crédit à une conception normative de nature épistémique. Une croyance est motivée lorsque les raisons de croire sont en fait des raisons de *vouloir* croire. Une croyance motivée prend typiquement la forme d'une pensée magique (*wishful-thinking*) qui consiste à prendre ses désirs pour des réalités, mais peut aussi prendre la forme d'un mensonge à soi-même (*self-deception*). Ce type de croyance vise à réduire ce que les psychologues appellent la dissonance cognitive. La dissonance cognitive apparait lorsque nous nous rendons compte que nous adoptons des croyances ou des lignes de conduite incohérentes. La dissonance cognitive apparait aussi lorsque nous devons vivre une existence normale alors que nous avons vécu un traumatisme important. Beaucoup de personnes minimisent l'impact d'un traumatisme pour réduire la dissonance cognitive dans le fonctionnement de leur esprit. La manière typique dont les croyances motivées permettent à certains agents de réduire leur souffrance psychique s'exprime dans le fait que l'impact des événements traumatisants est sous-estimé dans leur représentation mnémonique (Gilbert, 2006). Si notre seuil de tolérance à des dissonances est atteint, nous allons, suivant les situations, accepter des croyances parce qu'elles nous « conviennent » ou rejeter des croyances

Le fait de savoir si une bonne théorie de la faiblesse de volonté implique nécessairement une conception normative est aujourd'hui objet de débats. Même s'il est communément admis que la faiblesse de volonté est l'exemple paradigmatique de l'irrationalité pratique, beaucoup de philosophes estiment qu'on peut en fournir une conception purement descriptive. Aussi, ce qui est devenu la conception de référence de la faiblesse de volonté sur la scène philosophique actuelle, à savoir qu'elle consiste à agir à l'encontre de son meilleur jugement, est, avant interprétation, purement descriptive.

2.1.1.4 L'hypothèse de la maximisation du bien-être

Le cognitivisme motivationnel de Platon est en fait une conséquence intuitive de l'hypothèse que les agents tentent toujours de maximiser leur bien-être lorsqu'ils sont confrontés à des problèmes décisionnels. S'ils échouent à maximiser leur bien-être, la meilleure explication réside nécessairement dans le fait qu'ils ignoraient certains faits pertinents ou que certains événements extérieurs ont interféré avec le cours de leurs actions. L'hypothèse de la maximisation est clairement exprimée dans les passages des dialogues qui traitent des plaisirs et de leur mesure. La manière dont Platon envisageait la maximisation est tout à fait conforme à l'idée que s'en font les théoriciens modernes. : maximiser notre bien-être consiste à rechercher le meilleur résultat dans la balance complète des avantages et des inconvénients des options dont nous disposons pour résoudre un problème décisionnel.

Pour beaucoup de théoriciens modernes, l'hypothèse de la maximisation du bien-être est non seulement une hypothèse crédible ou probablement vraie, mais une caractéristique intrinsèque de la rationalité pratique. La plupart des axiomes des théories du choix rationnel en micro-économie reposent encore aujourd'hui sur l'idée que les agents cherchent toujours

---

parce qu'elles ne nous « conviennent » pas – dans le cas, par exemple, où elles sont incompatibles avec nos comportements habituels. Bien qu'elles jouent un rôle utile dans les cas de traumatismes psychologiques, l'adhésion à des croyances motivées contrevient à la norme épistémique qui stipule qu'on ne doit pas croire en quelque chose seulement ou principalement parce que cela nous « convient ».

La question est de savoir si, dans les cas de faiblesse de volonté, les croyances motivées motivent effectivement les actes de l'agent ou constituent seulement des justifications *a posteriori*. Mais, même en tant que justification, il y a tout lieu de croire que les croyances motivées jouent tout de même un rôle causal dans la fréquence des actes akratiques en ce sens qu'elles préparent le terrain pour des actes futurs et réduisent quelque peu la dissonance cognitive anticipée, laquelle constitue un frein à l'accomplissement d'acte akratique. Est-ce que tous les épisodes de faiblesse de volonté sont, ne serait-ce que de manière partielle, toujours motivés par des croyances motivées? Il est difficile de le dire, même si nos intuitions nous portent à croire que ce n'est pas le cas. Une chose est sûre cependant : les croyances motivées peuvent être extrêmement subtiles, et passent donc facilement inaperçues.

à maximiser leur bien-être – ou du moins leurs préférences (égoïstes ou altruistes). D'ailleurs, les théoriciens ne considèrent plus, du moins depuis Bentham, que le cognitivisme motivationnel soit une conséquence logique de l'hypothèse de maximisation. L'émergence des théories morales utilitaristes va de pair avec l'idée qu'il y a des éléments motivationnels de base de nature affective et désidérative (récompenses/punitions, plaisirs/déplaisirs, etc.). Aussi, la maximisation en est venue à porter exclusivement sur des éléments motivationnels beaucoup plus bruts (*raw*) que ceux qui ont à voir avec l'importance « réelle » qu'ils sont supposés avoir et identifiable par un exercice approprié de la raison.

L'existence putative d'épisode de faiblesse de volonté représente une difficulté pour toute théorie qui comporte l'hypothèse de maximisation. Une solution à saveur platonicienne consiste, comme on l'a vu, à jouer sur le paramètre cognitif. Les agents qui échouent à maximiser leurs intérêts bien pesés ignorent certains faits pertinents de manière ponctuelle ou persistante. Dans le cadre d'une théorie d'inspiration utilitariste, on pourra expliquer le phénomène en jouant sur les paramètres affectifs ou désidératifs. Si des agents, par ailleurs maximisateurs, adoptent des comportements manifestement sous-optimaux, c'est que leurs dispositions émotionnelles ou leurs désirs ont subi un changement. Si je remise, par exemple, mon appareil d'entrainement physique que j'ai pourtant acquis à grands frais il y a à peine un mois, c'est que je préfère maintenant m'avachir sur le canapé plutôt que de suer. Si je décide de prendre ma voiture avec un taux d'alcoolémie élevé, alors que je m'étais pourtant résolu à appeler un taxi plus tôt dans la soirée, c'est que l'alcool a réduit mon aversion et ma sensibilité globale au risque.

Bien que ce point semble être négligé dans la littérature sur la motivation, les théories de l'action et du choix rationnel, l'hypothèse de la maximisation est un des legs les plus importants (sinon le plus important) que nous ait laissé Platon. Qu'elle figure comme axiome explicite dans les théories normatives et descriptives en micro-économie, ou implicitement à la base des considérations intuitives pratiques des personnes, ou encore comme condition d'arrière-plan nécessaire à l'interprétation des actions d'autrui, l'hypothèse de la maximisation reste largement acceptée. D'ailleurs, c'est plus ou moins en référence à cette hypothèse que plusieurs philosophes ont nié l'existence de la stricte akrasie (Heath : 2008), et ce qui va s'imposer comme la conception standard moderne de la

faiblesse de volonté (« agir à l'encontre de son meilleur jugement ») y fait implicitement référence.

Malheureusement (ou heureusement), l'hypothèse de la maximisation est fausse. La description empirique des procédures décisionnelles depuis les premiers travaux d'Herbert Simon dans les années 1950 jusqu'aux récents travaux des membres du ABC Group de l'Institut Max Planck montrent très clairement que les agents sont des piètres maximisateurs, mais tirent leur épingle du jeu en utilisant toutes sortes de règles décisionnelles simples et efficaces (*fast and frugal heuristics*). Je vais présenter quelques-unes des conclusions les plus importantes de ces recherches au chapitre 6. Ces conclusions jettent, à mon avis, un éclairage nouveau et important sur les débats conceptuels à l'égard de la faiblesse de volonté.

2.1.2 Les défaillances du raisonnement pratique : les explications d'Aristote

Bien que le premier usage attesté du terme « *akrasia* » se trouve dans l'œuvre d'Aristote, il y est difficile de cerner une théorie ou un système explicatif complet de la faiblesse de volonté. Aristote aborde le problème à plusieurs endroits dans l'*Éthique à Nicomaque* – un traité qui porte sur les conditions de l'existence vertueuse et heureuse – et l'on peut également trouver des éléments de réponse dans *De Anima* – un traité sur la nature humaine et les principes de la vie en général. Comme je l'ai mentionné, plusieurs commentateurs soutiennent qu'Aristote interprétait la position de Platon comme une négation de l'existence de la stricte akrasie, et non de la faiblesse de volonté comme un phénomène plus général. Dans sa critique de la position platonicienne, il défend la possibilité de la stricte akrasie et offre des éléments pour une explication de celle-ci[13]. Aristote jugeait nécessaire de fournir de telles explications parce pour lui le cognitivisme motivationnel est contredit par nos intuitions les plus fortes (*Éthique à Nicomaque* : 1145b24-32). Par contre, certains commentateurs soutiennent que les explications fournies par Aristote ne lui permettent pas vraiment de faire l'économie du cognitivisme motivationnel.

[13] Je suis conscient qu'il faut être prudent quand on discute de la position d'Aristote sur ce genre de problème. Le Stagirite ne semble pas défendre un ensemble de thèses qu'ils jugent solidement assises, mais formule souvent des hypothèses provisoires qu'il tend à réviser ou à raffiner. Aussi, mon propos repose sur une sélection de certaines hypothèses d'Aristote que je juge les plus intéressantes pour expliquer le phénomène de la stricte akrasie.

Cela dit, Aristote voit les épisodes de faiblesse de volonté comme des manquements ou des errances ponctuelles qui ne peuvent s'inscrire dans aucun projet stable. Aussi, l'akratès n'est pas incontinent de nature, mais vit seulement des états transitoires où il exhibe une volonté faible (Gilbert Romeyer-Dherbey & Aubry, 2002). Ce sont des problèmes pratiques ponctuels qu'il faut expliquer en identifiant des causes ponctuelles. Or, Aristote ne propose pas une explication unique des épisodes de faiblesse de volonté, mais un cocktail explicatif – qui comprend par contre beaucoup d'éléments communs et qui se recoupent à certains endroits. Il propose trois explications de la faiblesse de volonté. Chacune met en relief une défaillance particulière du raisonnement pratique dont la fonction est d'orienter correctement l'agent dans sa quête du bonheur (*eudemonia*).

La première concerne l'ignorance de certains détails pertinents des actions. Elle semble au premier abord peu intéressante, mais elle met en relief un aspect de la faiblesse de volonté un peu négligé par les conceptions philosophiques contemporaines, et vaut par conséquent la peine qu'on s'y attarde un peu. La seconde explication met en relief le caractère impulsif de certains comportements akratiques. La troisième explication met en relief le rôle que jouent les émotions dans le processus de prise de décision chez les agents akratiques. Cette dernière explication est plus sophistiquée et a servi de référence pour toute la tradition gréco-latine et moderne. Mais elle est également conforme à certaines de nos intuitions communes sur les causes de la faiblesse de volonté. Aussi, j'en discuterai plus longuement.

1/ *L'ignorance de certains détails pertinents de nos actions*. Aristote identifie des cas d'errements pratiques où les agents ignorent certains faits (manières, conséquences, circonstances, principes d'action, etc.) relatifs à leur action. Il soutient que les agents ne sont pas susceptibles de blâme dans ces cas parce que leur action n'était pas volontaire. Si l'agent ignore que son action exemplifie certaines propriétés en vertu du fait qu'elle est accomplie de telle manière, qu'elle comporte telles conséquences ou est accomplie dans telle circonstance, l'agent est dit agir *dans l'ignorance*. À ce titre, l'agent est susceptible de provoquer l'indulgence, voire la pitié (*Éthique à Nicomaque* : 1109b30 et 1110b24). Par contre, il existe des cas où les agents commettent des errements pratiques, mais sont susceptibles de blâme. C'est le cas des agents qui agissent *par ignorance*. Ces agents ne se trouvent pas dans un état d'ignorance *malgré eux*, mais par négligence. Ces agents ont, pour ainsi dire, refusé de faire leurs devoirs épistémiques et pratiques.

Pour Aristote, le comportement de l'agent akratique est blâmable parce qu'il agit par ignorance. Il *devrait* connaitre certains faits pertinents concernant les lignes de conduite qui s'offrent à lui, mais se refuse en quelque sorte de le faire. Aristote mentionne même l'oubli d'une conclusion ou une d'étape antérieure du raisonnement pratique comme un facteur important de stricte akrasie. Pour lui, un agent qui vit un épisode de stricte akrasie oublie un jugement ou une résolution antérieure.

Cela dit, Aristote, en fin psychologue, distingue les cas où les agents connaissent certains faits pertinents pour prendre une bonne décision, mais ne les comprennent pas vraiment. Il est comme un étudiant capable de réciter par cœur des formules dont il ne mesure pas la portée. Aussi, un agent qui vit un épisode de stricte akrasie peut sembler connaitre les faits et les principes pertinents qui devraient l'orienter dans une autre direction, mais, par négligence, ne les comprend pas.

2/ *Le problème de la conclusion hâtive*. Aristote identifie également des cas d'errements pratiques où les agents connaissent et comprennent certains faits pertinents, mais sautent hâtivement à une conclusion qui les égare. Je peux estimer, par exemple, que toutes les boissons alcoolisées sont bonnes au goût et qu'il faut rechercher en gastronomie ce qui est bon au goût, et boire aussitôt un tonneau de bière. Ici, Aristote propose une explication de l'impulsivité typique de l'agent intempérant. Cette explication peut ressembler à la précédente. On est tenté de croire, dans le cas de l'intempérance gastronomico-alcoolique, que l'agent qui boit un tonneau ignore des faits pertinents concernant l'ingurgitation de grandes quantités d'alcool. Mais ce n'est pas nécessairement le cas. Un agent peut connaitre ces faits, mais n'en a pas tenu compte dans son raisonnement pratique. Il a conclu trop rapidement à partir du principe suivant lequel il faut rechercher ce qui est doux en gastronomie et de la croyance que l'alcool est un aliment doux.

Aristote explicite les étapes et la structure du raisonnement d'un agent à l'aide du syllogisme pratique. Un syllogisme pratique est un ensemble de prémisses qui en viennent à être mentalement connectées et produisent à ce moment une conclusion. Les prémisses expriment des buts ou des principes (*majeures*) et des moyens ou comptes rendus perceptuels (*mineures*) sur la base desquelles une conclusion est tirée par l'agent. La

conclusion d'un syllogisme n'a pas de contenu propositionnel, mais est l'action qui suit immédiatement la connexion des prémisses[14].

Aussi, pour reprendre notre exemple gastronomico-alcoolique, on pourrait dire que l'agent est allé trop vite et aurait dû inclure dans ses prémisses celle suivant laquelle il ne faut pas abuser des choses bonnes au goût (*majeure*). S'il avait pris en compte ce principe, et qu'il avait pu au préalable tout à fait le connaître et le comprendre, il n'aurait pas agi de la sorte. Le raisonnement pratique qui en résulterait déboucherait sur une conclusion différente, à savoir une consommation modérée d'alcool.

3/ *Le problème de la contamination affective du processus décisionnel*. La troisième explication fait appel à certaines caractéristiques générales de la mécanique du raisonnement pratique et au rôle délétère que peuvent y exercer à l'occasion les émotions. Tout ce processus s'enracine dans l'anthropologie aristotélicienne. Dans le livre II de *De Anima*, Aristote décrit les puissances et les fonctions de l'âme pour toutes formes d'organisme vivant. À côté des puissances végétatives, sensitives, locomotives, il y a les puissances appétitives et intellectives dont Aristote se sert pour expliquer la stricte akrasie et la faiblesse de volonté en général. Seuls les humains cumulent l'ensemble de ses puissances et sont pourvus d'intellect. Mais comme Aristote thématise l'akrasie comme une sorte de discorde entre les appétits et les forces intellectuelles, seuls les humains sont susceptibles d'akrasie.

Les appétits correspondent en fait aux émotions. Aristote distingue, au sein des appétits, ceux qui sont de nature sensible de ceux qui sont de nature intellective. Les appétits sensibles sont enracinés dans la partie sensible de l'âme. La faim, la soif, la douleur, l'angoisse, les frissons dus au froid et l'envie d'uriner sont des appétits commandés par notre nature sensible et se retrouvent chez la plupart des animaux. Pour Aristote, ces émotions représentent des forces purement irrationnelles qui peuvent nous mouvoir, mais de manière aveugle. Contrairement à ces derniers, les appétits intellectifs ont des contenus

[14] Il est contre-intuitif de stipuler que l'agent agit ou prend une décision aussitôt le processus de délibération achevé. On a l'impression qu'on peut obtenir un résultat à la suite d'une délibération et s'en servir plus tard, au moment où l'on jugera opportun de prendre une décision. La manière dont Davidson interprète Aristote sur ce point est intéressante. Le processus de délibération complété, l'agent agit ou prend une décision sur-le-champ, parce « *qu'il n'y a pas lieu de distinguer les conditions sous lesquelles l'agent est en position d'inférer qu'une action qu'il est libre d'accomplir est désirable des conditions sous lesquelles il agit* ». (1970 : 51-52).

cognitifs. Le désir de manger un burger, l'envie de boire du vin, la peur du voisin, par exemple, ont un contenu cognitif que n'ont pas leurs émotions brutes correspondantes, à savoir la faim, la soif et l'angoisse. Ce contenu est déterminé essentiellement par deux paramètres, soit l'intentionnalité (*son objet*) et l'antécédent interprétatif (*les croyances à l'égard de leur cause*)[15]. Ces paramètres cognitifs permettent notamment à Aristote de spécifier de manière plus précise la nature des émotions, qu'en mesurant les différences d'intensité et de durée. La colère, par exemple, est déclenchée par la croyance qu'on ait été l'objet d'un affront ou d'une humiliation imméritée et est orientée vers des individus particuliers, tandis que la haine peut s'orienter vers des classes d'individus comme des groupes ethniques ou des nations.

Si l'on résume la position d'Aristote avec un vocabulaire contemporain, on dira que les appétits sont soit des états somatiques, soit des désirs ou aversions qui ont un contenu propositionnel. Mais la satisfaction d'un désir ou d'une aversion a des répercussions somatiques qu'Aristote avait identifiées comme un apaisement de l'excitation[16], et que la neurologie actuelle identifie comme un retour à un état homéostatique stable.

Les forces intellectuelles sont rationnelles et se divisent également en deux systèmes. Il y a l'intellect contemplatif, qui correspond à la raison théorique. Il concerne les raisons de croire et les règles *épistémiques* qui encadrent l'inférence de ces croyances. Et il y a l'intellect pratique, qui correspond à la raison pratique. Il concerne les raisons de faire et les règles *prudentielles* qui encadrent l'accomplissement d'actions. Aristote appelle *prudence* la disposition générale à utiliser des règles pour s'orienter dans notre quête du bien et de l'évitement du mal (*Éthique à Nicomaque* : 1140b 21). Se montrer prudent est, pour Aristote, l'exercice par excellence de la rationalité pratique. Mais la prudence ne porte pas seulement sur les moyens pour atteindre nos fins. Elle fixe les buts qui valent la peine pour

[15] Pour Jon Elster, la mise en relief des antécédents cognitifs (*interprétatifs*) par Aristote est une des avancées majeures dans l'étude des émotions, mais qui ne fut malheureusement pas suffisamment prise en compte par la tradition philosophique et psychologique (1999 : 118-119).

[16] Aristote affirmait que le physicien et le dialecticien peuvent tous deux parler des affections du corps, mais en des termes différents. Le physicien définirait la colère comme une ébullition du sang ou d'une substance chaude entourant le cœur, cependant que le dialecticien la définira comme l'appétit de rendre la souffrance pour la souffrance, etc. (*De Anima* : 403a17 29-32).

l'agent de poursuivre et détermine certains de ceux qui n'en valent pas la peine ou que l'agent doit éviter[17].

Aussi, pour Aristote, les finalités établies par la raison pratique de l'agent prudent constituent les points de références ultimes pour évaluer les appétits de niveau inférieur ou même déterminer leur contenu cognitif[18]. Ces finalités servent de point de départ normatif du raisonnement pratique et filtrent en quelque sorte les appétits de plus bas niveau. Mais lorsqu'un appétit force les mailles du filtre ou le contourne, il motive plus ou moins directement l'agent à s'engager dans une ligne de conduite incontinente. La satisfaction des appétits de l'agent incontinent l'éloigne de son bonheur par une contamination de son raisonnement pratique.

Par exemple, si je dois rechercher le bonheur, alors mon envie de consommer de l'alcool jusqu'à plus soif ne devra pas être considérée dans une délibération pratique parce que satisfaire cette envie engendre des problèmes de santé qui m'éloigneraient du bonheur. Or, je le considère de manière plus ou moins irrésistible et cela me motive à abuser de l'alcool.

Mais Aristote ne bannit pas pour autant les émotions (appétits sensibles et intellectifs) de la quête du bonheur. Les émotions *peuvent* jouer un certain rôle dans cette quête. Les émotions doivent être compatibles avec les objectifs principaux qui sont fixés par la raison pratique, pour deux raisons. La première est qu'il est préférable que les agents aient du plaisir à poursuivre leurs objectifs (*Éthique à Nicomaque* : 1105a 14). La seconde, plus importante pour notre propos, est que les émotions semblent jouer pour Aristote un rôle motivationnel que ne peut pas jouer ou peut difficilement jouer la raison pratique. La raison pratique fixe de manière prudentielle les objectifs généraux et spécifie les moyens d'y parvenir, tandis que les émotions portent plutôt sur des détails, ce qui motive souvent les agents à poursuivre des objectifs ponctuels et superficiels. Or, Aristote soutient qu'un agent

[17] Les considérations prudentielles sur les fins sont, chez Aristote, de nature essentiellement morale : « *la vertu est ce qui fait que nos objectifs sont corrects, et la sagesse pratique nous y amène* » (*Éthique à Nicomaque* : 1144a 7-8).

[18] Par exemple, je peux interpréter un état somatique de faim comme l'envie de manger une pomme et non comme l'envie de manger un baril de poulet frit *parce que* j'ai le projet d'acquérir de bonnes habitudes alimentaires, lesquelles sont commandées par des règles prudentielles de poursuite du bonheur. Mais les considérations prudentielles ne sont évidemment pas suffisantes ici pour isoler le contenu d'un appétit. On a besoin en plus de considérations factuelles supplémentaires, sans quoi on pourrait s'égarer dans l'interprétation de ces appétits. Je peux, par exemple, interpréter à tort un état somatique engendré par la fréquentation d'une personne comme de l'admiration, alors qu'il ne s'agit que d'une vile envie d'obtenir ses richesses. Pourtant, la vile envie peut rester tout à fait incompatible avec mon projet de vie vertueuse.

peut devenir akratique parce qu'il échoue à concrétiser ses raisons de niveau supérieur dans des raisons de niveau inférieur portant sur des détails tels les émotions[19].

Aussi, les émotions (ou appétits) ne sont pas que des contaminants possibles qui doivent être filtrés par la raison pratique, mais jouent un rôle essentiel pour régler les détails de nos vies. C'est pour cela qu'il est important de bien les filtrer, et de les intégrer à notre bagage motivationnel sous-tendant de saines habitudes de vie. Un agent qui a acquis de saines habitudes concernant les tâches ménagères sera moins susceptible de les repousser de manière irrationnelle parce qu'il aura plus souvent envie de les faire dans les délais.

2.1.2.1 Manquements épistémiques et impulsivité

La première explication d'Aristote (le facteur ignorance) est intéressante à plusieurs égards. Elle met en relief des conditions cognitives nécessaires au diagnostic de faiblesse de volonté. Si ces conditions minimales ne sont pas remplies, il ne peut pas diagnostiquer un épisode de faiblesse de volonté. Par exemple, il y a une différence majeure entre l'attitude des fumeurs des années 1950 et celles des fumeurs des années 2000 – du moins en Occident. Les premiers ne possédaient pas le bagage informationnel pour estimer adéquatement le degré de nocivité. Aussi, on ne peut considérer que leur consommation relevait de la faiblesse de volonté. Les choses sont différentes pour les seconds.

D'ailleurs, la distinction entre agir *dans* l'ignorance et *par* ignorance est utile et n'est à mon avis pas suffisamment exploitée dans la littérature contemporaine sur la faiblesse de volonté. Elle met relief les attitudes négligentes dans l'explication de certains épisodes de faiblesse de volonté. Je peux, par négligence, ne pas m'informer suffisamment sur les conséquences de la consommation d'alcool sur le fœtus et boire par conséquent davantage que ce qui est recommandé par les autorités médicales. Il n'est pas contre-intuitif de supposer que ma volonté est faible dans ce cas si j'ai délibérément négligé l'examen antérieur d'informations pertinentes qui pourtant sont disponibles. Mais on pourra argumenter que je n'ai pas vécu un épisode de faiblesse de volonté au moment où j'ai décidé de boire trop d'alcool, mais plutôt au moment où j'ai décidé de ne pas m'informer davantage. On peut toutefois estimer qu'Aristote a raison de considérer la négligence

[19] Aristote parle d'un échec de concrétisation d'une idée universelle dans l'idée du particulier comme d'une défaillance supplémentaire du raisonnement pratique. (*Éthique à Nicomaque* : 1102b 28-30)

épistémique comme un des facteurs de faiblesse de volonté (en général), sans qu'elle soit un facteur de stricte akrasie. Ce dernier point est important parce qu'il met en relief des cas de faiblesse de volonté qui échappent aux analyses actuelles basées sur la planification rationnelle (Bratman, 1987 ; Holton, 1999, 2009). Aussi, j'en dirai davantage au chapitre consacré à l'exposition de ces analyses (Chapitre 7).

Cela dit, Aristote, en distinguant la compréhension véritable d'une connaissance si superficielle qu'elle équivaut en gros à de l'ignorance, anticipe, d'une certaine manière, les critiques sceptiques contemporaines sur la possibilité de la stricte akrasie[20] L'insincérité, la mauvaise foi, l'hypocrisie, le bon ton, et quelques formes de duperie de soi reposent souvent sur cette forme d'ignorance et expliqueraient beaucoup de cas de faiblesse de volonté (Davidson, 1970). J'aborderai ces arguments sceptiques dans le chapitre 5.

La seconde explication d'Aristote est également intéressante. Il ne fait pas de doute que nombre d'épisodes de faiblesse de volonté sont motivés par des conclusions hâtives. La négligence peut également y jouer un rôle, comme dans l'explication précédente[21]. Négligeant de faire mes devoirs épistémiques, je ne pousse pas ma délibération assez loin et tire par conséquent une conclusion hâtive sur ce que je dois faire. Mais les émotions peuvent jouer un rôle analogue. Sous le coup d'une émotion forte, je peux abréger ma délibération et tirer une conclusion hâtive. Ce type d'explication vise le caractère impulsif de certains épisodes de faiblesse de volonté.

Cela s'accorde avec l'étiologie de bon nombre d'épisodes de faiblesse de volonté. Les *déclencheurs* et les *motivations* viscérales sont des causes proximales de beaucoup d'épisodes de faiblesse de volonté et affectent le processus de délibération rationnelle, un peu de la manière dont Aristote le conçoit ici.

Les *déclencheurs* peuvent être des stimuli délétères, qui causent des épisodes de faiblesse de volonté quand ils entrent dans le champ perceptuel de l'agent. Voir un désert, des amis fumer, une seringue, la télévision, un poste d'ordinateur, des biens de luxe, un casino ou une machine à sous, peut causer la rechute d'un agent qui avait de saines résolutions. Ce type de déclencheur obscurcit brièvement certaines raisons, au point qu'elles sont ignorées

[20] Même si ce n'est évidemment pas dans cette direction que voulait aller Aristote.
[21] Ici, la seconde explication recoupe la première.

ou qu'elles ne sont plus suffisamment prises en compte, lors d'une nouvelle délibération. Mais les déclencheurs ne sont pas tous de nature perceptuelle. Apprendre qu'une certaine option est maintenant disponible dans mon environnement, sans l'observer directement, est suffisant, dans certains cas, pour engendrer un épisode de faiblesse de volonté. Joueur pathologique repenti, je peux vivre une rechute seulement en apprenant qu'un casino vient d'ouvrir ses portes dans la ville où j'habite. Dans ce genre de cas, je n'ai pas de contact direct avec un stimulus délétère. Le simple fait d'apprendre ou de savoir qu'une certaine option est maintenant disponible dans mon environnement est un facteur de faiblesse de volonté. Beaucoup de salariés, par exemple, prévoient payer leurs dettes aussitôt qu'ils obtiendront leur paie, mais changent d'avis aussitôt qu'ils l'ont effectivement obtenue. À ce moment, beaucoup d'options d'achat deviennent disponibles et cela motive des comportements dépensiers. D'ailleurs, les déclencheurs ont un effet émotionnel important. Nous avons plus envie d'uriner lorsque nous voyons une toilette, même s'il n'y a pas un volume d'urine plus grand dans notre vessie à ce moment. Nous avons davantage envie de fumer dans les lieux où nous avons le droit que dans les lieux où c'est rigoureusement proscrit.

Contrairement aux déclencheurs, les *motivations viscérales* sont des facteurs endogènes de changement des préférences (Elster, 2007 ; Loewenstein & Preleck, 1993). Elles peuvent à l'occasion pousser les agents à accomplir des actions qu'ils regretteront ensuite. Elles consistent *grosso modo* à une émotion forte causée par le *manque* ou le *surplus* d'une substance ou d'une quantité calorique dans l'organisme de l'agent. Lorsque le niveau de ces éléments dans l'économie de l'organisme atteint un certain seuil (inférieur ou supérieur), il enclenche un état émotif qui capte l'attention de l'agent et l'oriente dans une nouvelle direction comportementale. Les états d'ivresse engendrés par un niveau d'alcoolémie trop grand rendent les agents plus disposés à prendre des risques importants, dont prendre leur voiture avec des facultés affaiblies. Sous l'influence de l'alcool, notre délibération est de mauvaise qualité parce que bon nombre d'aspects pertinents de la situation ne sont pas ou sont insuffisamment pris en compte. Les états de manque des consommateurs de drogue, la soif excessive qui motive les naufragés à boire de l'eau de mer ou leur urine, la faim véritable par opposition à la gourmandise, l'envie d'uriner et de déféquer sont tous des

changements corporels qui résultent d'un processus initié à l'intérieur du corps et non à la faveur d'un stimulus.

Les déclencheurs et les motivations viscérales expliquent beaucoup de cas d'impulsivité. Ces causes obscurcissent certaines raisons pourtant pertinentes et poussent les agents à tirer trop rapidement des conclusions. Qu'on accepte ou non ces éléments de classification des causes de l'impulsivité, on peut cependant critiquer la position d'Aristote sur un point. Il n'est pas toujours irrationnel d'agir de manière impulsive. Or, comme la faiblesse de volonté est le paradigme de l'action irrationnelle, il s'ensuit que l'analyse d'Aristote est ici, dans le meilleur des cas, incomplète. Il lui faudrait expliquer pourquoi certains comportements impulsifs relèvent de la faiblesse de volonté alors que d'autres non.

Comme délibérer prend du temps, l'urgence d'une situation peut commander une action sur le champ. Aussi, la perception d'une urgence crée en nous une émotion qui nous pousse à tirer une conclusion sans avoir au préalable examiné tous les tenants et aboutissants. Sans court-circuiter pour autant nos processus de délibérations, agir sous le coup d'une émotion forte peut consister *à contourner* complètement ces processus en vue d'une réponse plus rapide à une situation problématique, ou perçue comme problématique. Il y a ici un avantage adaptatif indubitable à réagir impulsivement dans certaines situations. Il est préférable de surestimer que sous-estimer des dangers potentiellement mortels. Par exemple, je peux, après avoir confondu brièvement une branche d'arbre avec un serpent, déguerpir sans prendre le temps d'interpréter correctement le stimulus. Il est préférable de déguerpir quand j'ai affaire à une branche qui ressemble à un serpent que de m'en abstenir quand j'ai affaire à un serpent qui ressemble à une branche. Bien entendu, il est préférable que je ne déguerpisse *que* lorsque le danger est réel. Pour m'assurer que le danger est bien réel, je dois délibérer et tenir compte de plus d'information. Or, les processus de délibération prennent du temps, et il est bon de s'y engager lorsque cela augmente les chances de succès ou diminue les risques d'échec. De sorte qu'une délibération *peut* minimiser les chances de succès ou maximiser les risques d'échec si le temps de décision est un paramètre important. C'est notamment ce qui se passe dans les situations

d'urgence[22]. Dans ce genre de situation, il est plus rationnel de s'en abstenir même si cela débouche à l'occasion sur de fausses alertes. Le problème pour l'agent rationnel n'est pas de prendre une décision impulsive quand une alerte sonne, mais de savoir si les conditions de l'alerte fiable sont remplies. Si je me trouve au Québec, par exemple, ma réaction impulsive devant une branche sera surprenante. Les serpents qu'on trouve dans les forêts québécoises ne sont que d'inoffensives couleuvres. Ma réaction sera par contre compréhensible si j'ai passé quelques années en Arizona ou dans les Everglades. Je sais toutefois que je devrai m'adapter à l'environnement naturel québécois de manière à ce que les alarmes cessent de retentir chaque fois que je risque de mettre le pied sur une branche qui ressemble à un serpent.

2.1.2.2 L'intrusion des émotions dans les « champs de compétence » de la raison

La troisième explication d'Aristote est sans contredit la plus intéressante des trois parce qu'elle vise à rendre compte non seulement de la faiblesse de volonté en général, mais de la stricte akrasie en particulier. L'incompatibilité des lignes de conduite qui sont motivées par des raisons de niveau supérieur (issues de la raison ou de l'intellect) et des raisons de niveau inférieur (issues des systèmes appétitifs ou émotionnels) résume le conflit intérieur typique des agents akratiques. Lorsque nous nous apprêtons à prendre une seconde portion de dessert alors qu'il y a une petite voix intérieure qui nous enjoint de nous en abstenir, nous mettons en concurrence des raisons et des préférences de niveaux différents.

D'ailleurs, nous avons des intuitions relativement claires sur ce qui distingue ces niveaux et n'avons pas besoin de connaître la théorie aristotélicienne pour le mettre en relief. Les raisons de niveau supérieur entrent souvent dans la formation de préférences *a*temporelles et souvent impersonnelles (objectives) comme « on ne devrait pas fumer », « on devrait toujours tenir ses promesses » ou « il est préférable de faire ce que les professeurs nous demandent! », etc. Tandis que les raisons de plus bas niveau entrent dans la formation de préférences indexées au temps et plus personnelles (subjectives) comme « j'ai envie de fumer tout de suite », « je préfère pour cette fois seulement ne pas tenir ma promesse » ou « je n'ai pas envie de faire mes devoirs ce soir même si le professeur me l'a demandé ».

[22] Pour être plus précis, c'est la structure du problème décisionnel qui commande ou non une délibération rationnelle complète, et dont l'échéancier n'est qu'un aspect parmi d'autres. Je développerai ce point plus en détail au chapitre 6.

Aussi, les premières préférences correspondent typiquement à la voix de la raison, alors que les secondes correspondent plutôt à la voix de la passion. C'est-à-dire que nous avons l'intuition que les raisons qui entrent dans la formation des premières sont dénuées d'émotion, alors que les secondes sont clairement noyautées (contaminées?) par les émotions.

Or, nous ne sommes pas seulement enclins à penser ici qu'il y a une différence au niveau de la structure et des antécédents motivationnels et justificationnels. Nous estimons spontanément que nous devons accorder notre conduite avec les raisons de niveau supérieur parce qu'elles sont de *meilleures raisons* et qu'il est, par conséquent, plus rationnel de le faire. Elles sont de meilleures raisons parce qu'elles relèvent de considérations variées, mais pertinentes, et constituent un arrière-plan cognitif plus riche, que l'arrière-plan cognitif constitué par des émotions. Les raisonnements pratiques « passionnés » sont cognitivement plus pauvres que les raisonnements « dépassionnés »[23]. C'est sans doute pour cette raison qu'on estime intuitivement dans beaucoup de situations que moins de raisons militent en faveur de choix par trop enthousiastes que de choix plus circonspects.

En dépit de son raffinement supérieur, la position d'Aristote à l'égard des émotions reste sensiblement la même que celle de Platon. Les considérations affectives ne doivent motiver les décisions que dans la mesure où elles s'accordent avec les orientations générales de la froide raison. Elles facilitent la réalisation d'actions rationnelles dans le meilleur des cas, mais l'entravent dans beaucoup d'autres. Aussi, pour Aristote, les épisodes de faiblesse de volonté sont largement explicables par l'intrusion d'émotions inappropriées dans le raisonnement pratique. Sans adhérer pour autant au cognitivisme motivationnel, Aristote reste conforme à certaines propositions platoniciennes, notamment celles qui confèrent à la raison la tâche d'attribuer des poids aux diverses motivations en fonction de leur importance relative réelle. À ce titre, la conception aristotélicienne de la motivation prête le flanc aux mêmes critiques que celles de Platon : assigner des poids à des désirs et à des émotions est également déterminé par des poids issus de désirs et d'émotions.

[23] Même si un raisonnement dépassionné peut arriver au même résultat qu'un raisonnement passionné. Je peux avoir des raisons variées et pertinentes d'accepter une loi en faveur de la peine de mort ou de ressentir de la colère suscitée par l'examen sommaire d'un cas d'homicide volontaire. Par contre, comme le fait remarquer Elster, il semble y avoir des émotions plus compatibles que d'autres avec le calcul rationnel. La colère obscurcie plus que la haine le processus de décision. La colère nous pousse à prendre des risques importants et à agir à l'encontre de nos intérêts. La haine repose seulement sur des prémisses erronées (Politique : 1315a 25-31).

Il se distingue toutefois de Platon en acceptant la réalité de la stricte akrasie et en l'expliquant par l'intrusion de certaines émotions dans les « champs de compétences » de la raison au moment de la réalisation du raisonnement pratique. Pour utiliser une métaphore politique, un agent akratique serait comme un gouvernement qui décide pour des raisons bassement électoralistes d'utiliser la politique monétaire à son profit en passant par-dessus la tête de la Banque Centrale (ou Réserve Fédérale) en dépit du fait qu'en vertu de la Constitution du pays, il n'est pas autorisé à le faire, mais qui dispose néanmoins de l'assentiment d'une masse critique de citoyens ce qui lui confère suffisamment de pouvoir sur le terrain.

Certains ont critiqué la position d'Aristote sur la stricte akrasie en soutenant que les facteurs affectifs ne semblent pas jouer un rôle dans bon nombre de comportements akratiques, comme succomber à des tentations ou se montrer immodéré dans la consommation de biens. Suivant en cela Austin (1961), Davidson part du constat que nous succombons bien souvent à la tentation « avec calme et même avec subtilité » (1970 : 47-48). Il s'ensuit qu'on ne peut pas expliquer les comportements relevant de la stricte akrasie par des soubresauts émotionnels, par des penchants désirants, par des convoitises, par des aspirations exaltantes, ou même par quelques autres émotions cependant plus tranquilles et lancinantes. Nous pouvons, semble-t-il, vivre des épisodes de faiblesse de volonté alors que nous sommes dans des états dépourvus d'émotions[24].

Cette critique d'Aristote n'est cependant pas aussi solide qu'il y paraît. On a au moins deux raisons de penser que la critique de Davidson n'est pas concluante. La première porte sur la fiabilité des rapports conscients que les agents font à l'égard de la réalisation en eux d'une émotion. La seconde porte sur l'apport des émotions dans l'orientation de l'attention.

1/ Les rapports en première personne d'une expérience émotionnelle peuvent faire l'objet de plusieurs erreurs. La plus recensée dans la littérature de psychologie expérimentale est l'erreur d'attribution causale et d'interprétation (Nisbett & Wilson, 1977 ; Murphy & Zajonc, 1991). Mais l'état des recherches récentes en psychologie et en neurologie atteste

[24] Jon Elster critique également la théorie émotiviste de l'akrasie. Selon lui, le phénomène de la dynamique des préférences temporelles et l'escompte hyperbolique du futur sont des contre-exemples suffisamment convaincants pour discréditer cette théorie, du moins dans sa prétention de théorie générale. Je discuterai en détail de ces notions dans le dernier chapitre.

également l'existence d'émotions inconscientes. Cela contredit la conception classique des émotions depuis Williams James. Pour James, les émotions sont des rétroactions sensorielles de réactions involontaires à des événements perçus par un agent. Aussi, la conscience de la qualité (plaisante ou déplaisante) d'une émotion n'est pas un trait accessoire de l'émotion, mais une de ses caractéristiques essentielles[25]). Or, Zajonc (1991), Kihlstom (1999), Berridge et Winkielman (2003) ont mesuré et documenté des expériences affectives marquées chez des sujets pourtant incapables d'identifier ou même de reconnaître qu'ils vivent une émotion au moment où elle est causée. Pour ces chercheurs, les rapports en première personne ne sont qu'une source relativement peu fiable d'information à l'égard de la condition émotive d'un sujet. L'observation détaillée des expressions faciales involontaires, du tonus musculaire, des mouvements de la langue, des choix de consommation, de la conductance de la peau, et diverses méthodes de scanographie du cerveau, constitue des sources d'information beaucoup plus fiables. Les chercheurs ont mis les sujets dans des situations contrôlées qui comportaient des stimuli qui passaient la plupart du temps, mais pas tout le temps, sous le radar de la conscience. Il est particulièrement facile de déclencher des réactions émotives inconscientes à l'aide de stimuli inconscients ou partiellement inconscients (image subliminale, désinformations, orientation de l'attention, etc.). Damasio propose, quant à lui, de distinguer les émotions – qu'il conçoit comme un programme réactionnel déclenché par un stimulus externe ou interne – des sentiments d'émotion. Ces derniers seraient des sortes de cartes mentales portant sur l'état du corps ou de certaines de ses parties au moment où une émotion est réalisée (Damasio, 1995, 2010). Les sentiments d'émotion ne sont pas une condition nécessaire pour la réalisation d'un programme émotionnel, ce qui laisse place à la possibilité d'émotions inconscientes (Berridge & Winkielman, 2003). D'ailleurs, nous estimons intuitivement que les nourrissons et les mammifères éprouvent des émotions simples sans pour autant en avoir conscience, et il nous est tous déjà arrivé d'être le témoin abasourdi de cris proférés par une personne qui niait de manière colérique être en colère! Ce qu'on peut inférer de ce qui précède est qu'on ne peut pas critiquer la position d'Aristote au sujet des causes émotionnelles en ne se basant que sur nos propres rapports

[25] James résuma dans une phrase célèbre sa conception des émotions : «*[…] that our feeling of the same changes as they occur IS the emotion* » (1884 : 199).

concernant nos états émotionnels pendant un épisode de faiblesse de volonté. Peut-être qu'Aristote a tort de pointer les émotions comme cause principale des comportements de stricte akrasie comme le fait de succomber à une tentation ou de surconsommer. Mais nos rapports en première personne ne sont pas une source d'information suffisamment fiable pour trancher la question.

2/ L'autre raison réside dans l'hypothèse – plus populaire – que les émotions semblent faire partie intégrante de tous les processus décisionnels qui portent sur des options qui ont une importance particulière pour l'agent (choix égocentriques) ou qui comporte des conséquences hédoniques significatives pour lui (choix intéressés)[26]. De Sousa suggère que même dans les situations où les agents akratiques succombent à la tentation « à froid », les mécanismes de l'attention sont en partie déterminés par des réactions émotionnelles (1987 : 243-244). Selon Ainslie, les intérêts à court terme que favorisent les actes akratiques présentent une structure dynamique de récompenses qui gère l'attention de l'agent. Cela explique pourquoi il est souvent si difficile de maintenir le cap dans l'exécution de plan et de résolution à long terme (2001 : 76-77). Si les émotions sont un « input » important des mécanismes d'attention, il y a tout lieu de croire qu'elles jouent un rôle, ne serait-ce qu'indirectement, dans les choix akratiques.

Cela dit, on peut accepter l'idée d'Aristote de faire des considérations affectives un des principaux facteurs de faiblesse de volonté. Seulement, la conception du rôle des émotions (et des désirs de plus bas niveau) dans le processus décisionnel qu'il promeut est aujourd'hui largement contestée.

Aristote se fit l'apôtre – comme Platon et beaucoup d'autres penseurs de son époque – d'un modèle de la rationalité qu'on pourrait qualifier de « *vulcain* », et c'est ce qui explique pourquoi il fut si prompt à bannir les émotions. Pour ceux qui ne connaissent pas la série télévisée *Star Trek*, Spock, le fidèle compagnon de route du capitaine du vaisseau USS Enterprise, James Kirk, appartient au peuple qui habitait – avant sa destruction – la planète Vulcain. Les vulcains ont une caractéristique motivationnelle et cognitive singulière : comme ils ne ressentent pas d'émotions, ils peuvent analyser, à la lumière des principes les plus froids de la rationalité, chaque situation avant de prendre une décision. Ce modèle de

[26] Damasio, *Ibid*.

rationalité est largement répandu dans la population. La plupart des personnes estiment que prendre des décisions rationnelles requiert un état mental dépourvu d'émotions. Si Spock fut le personnage le plus populaire de la série, c'est probablement en partie parce que les téléspectateurs partagent le préjugé qui veut que les émotions soient incompatibles avec les canons de la rationalité. Seulement, on observe dans plusieurs épisodes de la série, un Spock qui se montre incapable d'identifier le choix le plus adéquat dans des situations complexes où les considérations morales sont importantes et hautement chargées affectivement. Les téléspectateurs ont l'impression que la raison froide du personnage n'est pas « éclairée » par un fond émotionnel et qu'elle reste en quelque sorte aveugle à un niveau de réalité pourtant important.

Cette dernière opinion s'accorde avec nos intuitions morales et esthétiques fondamentales. Si nous ne sommes pas touchés par la situation déplorable d'une personne sur laquelle le sort s'acharne, ou par les conditions de vie misérables des enfants affamés du tiers-monde, ou par l'inventivité présente dans un chef-d'œuvre, nous avons l'impression que quelque chose nous échappe ou que nous ne comprenons pas bien la situation. Faire des expériences nouvelles pour étendre notre sensibilité est sans doute un bon moyen d'acquérir des « connaissances » auxquelles nous avons difficilement accès en nous contentant d'informations discursives, comme celles qu'on retrouve dans les livres, les articles et les journaux[27]. Je ne remets pas en cause ici l'idéal d'objectivité que les philosophes chérissent depuis l'Antiquité. En fait, cet idéal vaut non seulement la peine d'être poursuivi, mais est même accessible lorsque correctement conçu. Il devient inaccessible et même nuisible lors qu'on assimile cet idéal avec le point de vue cosmique, ou le point de vue de nulle part. Ce point de vue, tel que décrit par Nagel dans un ouvrage devenu classique, requiert que nous nous détachions de nos intérêts particuliers, dispositions émotionnelles et même de la structure fondamentale de notre expérience perceptuelle – en particulier de sa relativité et son idiosyncrasie – bref, de toutes perspectives particulières ou humaines[28]. Cela ne

[27] Loewenstein et Small ont même découvert que l'attrait du comportement charitable à l'égard d'une seule victime à laquelle on s'identifie diminue lorsque lors qu'on nous fournit des informations statistiques pertinentes au cas (2007 **:** 112-126).

[28] Le souci d'une objectivité détachée qui se rapproche du point de vue cosmique se trouve exprimé chez Aristote dans ce qu'on appelle la version non scolaire du syllogisme pratique. Un syllogisme pratique est l'articulation verbale des inférences que les agents font lorsqu'ils délibèrent en vue d'accomplir une action. Dans sa version « scolaire » un syllogisme est constitué d'une prémisse qui identifie un but (la majeure) et une prémisse qui identifie un moyen pour atteindre ce but (la mineure). Suivant les interprétations, la conclusion indique un choix d'action ou une ligne de conduite

concorde toutefois pas avec la conception « ordinaire » de l'objectivité. Lorsque nous remettons en cause le jugement d'une personne parce que nous estimons qu'elle n'est pas suffisamment objective, c'est que nous estimons qu'elle aurait dû faire abstraction ou du moins prendre du recul à l'égard de *certains* de ses intérêts, préférences ou dispositions émotionnelles particulières, et non pas de tous[29].

En fait, les émotions doivent même être adaptées aux situations. L'indignation est plus appropriée que la colère lorsque nous voyons un enfant mourir de faim. La colère est plus appropriée lorsque nous apercevons un enfant se faire battre à mort. Ces différentes émotions nous motivent à faire des choses fort différentes : militer pour conscientiser les gens à donner davantage à des œuvres de charité, d'une part, nous montrer prêts à nous interposer avec violence, d'autre part.

On pourrait tout de même continuer à défendre quelque chose qui ressemble au modèle vulcain de la rationalité pratique en soutenant que les émotions peuvent être utiles pour prendre des décisions rapides dans des situations qui l'exigent, mais que l'usage de la raison froide constitue l'étalon de mesure qui permet *en dernier ressort* d'évaluer correctement des décisions. Cela reste compatible avec l'idée que les émotions ont une fonction adaptative particulière : un système d'appoint pour les fonctions cognitives plus sophistiquées. Pour reprendre un exemple qu'Elster a proposé dans *Alchimie of Mind*, je peux faire une promenade dans la forêt et apercevoir sur mon chemin quelque chose qui ressemble à un serpent et, sous le coup de la peur, décider sur le champ de m'enfuir dans la direction opposée. Si ce que j'ai pris pour un serpent s'avère être en fait une branche, j'ai alors couru pour rien. Mais il est préférable que je me sois trompé en prenant une branche pour un serpent qu'un serpent pour une branche. Le processus de prise de décision est plus

---

particulière, ou est elle-même cette action ou cette ligne de conduite. Par exemple, me demandant comment je dois m'y prendre pour me rendre à un rendez-vous, je peux faire l'inférence suivante : je souhaite arriver à mon rendez-vous à l'heure (prémisse majeure), prendre ma voiture est un bon moyen pour arriver à mon rendez-vous à l'heure (prémisse mineure), je prends donc ma voiture (conclusion). La version scolaire du syllogisme exprime clairement une inférence inductive suivant le schéma fin/moyen. Par contre, on trouve dans l'œuvre d'Aristote une version différente du syllogisme pratique. Il exprime une inférence déductive suivant le schéma universel/particulier. La majeure est censée exprimer un bien universellement désirable, tandis que la mineure un fait particulier pertinent pour déterminer l'action à accomplir. Si je reprends mon exemple, j'obtiens quelque chose comme : il est bien d'arriver à l'heure à un rendez-vous (majeure), il est bien d'utiliser un véhicule si cela me permet d'arriver à l'heure d'un rendez-vous (majeure), prendre ma voiture permet d'arriver à l'heure de mon rendez-vous (mineure), ceci est ma voiture (mineure), je prends donc ma voiture (conclusion). Cette version du syllogisme pratique apparait particulièrement artificielle. Les désirs des agents sont traités dans des formules qui minimisent voire font disparaître leur idiosyncrasie, ce qui laisse croire qu'on gagne en objectivité.

[29] L'erreur philosophique qui consiste à assimiler l'objectivité au point de vue cosmique relèverait peut-être d'une simple généralisation hâtive (Nathanson : 1994).

rapide lorsqu'il emprunte des circuits émotionnels que lorsqu'il emprunte des circuits délibératifs[30]. Des émotions fortes comme la peur nous font typiquement préférer une action immédiate à une action différée. Or, l'action immédiate présente dans certaines situations un avantage adaptatif indéniable. À défaut d'utiliser l'artillerie lourde de la délibération rationnelle, mieux vaut se rabattre sur l'émotion. Mais lorsque nous disposons de suffisamment de temps, il est plus judicieux de sortir les gros canons. C'est toujours la délibération rationnelle et l'usage de nos capacités cognitives les plus sophistiquées qui, ultimement, nous permettent d'obtenir le fin mot de l'histoire. La rationalité froide nous permet de distinguer les faux positifs des faux négatifs, mais il reste statistiquement avantageux dans des situations où nous devons réagir rapidement à des sources potentielles de dangers d'orienter notre action en fonction de faux négatifs que de faux positifs.

Cet argument permet-il de préserver le modèle vulcain? En un sens, oui. Mais dans une version plus faible. Un agent rationnel pourrait se guider en fonction d'émotions, même fortes, dans certaines situations d'urgence, mais il devra évaluer *post facto* son choix par l'exercice de sa froide raison. Cette version correspond à ce que Bernard Gert a appelé la « *cool moment theory* », une hypothèse normative pour évaluer les désirs en général – et pas seulement ceux motivant la prise de décision en situation d'urgence (1988 : 42-44). La théorie du « *cool moment* » stipule qu'un désir est irrationnel si et seulement si sa satisfaction implique la frustration de désirs qui, dans un moment de délibération froide, sont *jugés plus importants*. Aussi, s'il est irrationnel d'adopter des comportements akratiques, c'est que ces comportements satisfont des désirs qui sont jugés dans des moments froids moins importants que les désirs dont ils impliquent la frustration. Gert critique cette hypothèse normative parce que, selon lui, elle ne permet pas de considérer comme irrationnels certains désirs qui sont pourtant clairement irrationnels comme le souhait de devenir esclave ou de se couper un membre parce que ces désirs pourraient pour certaines personnes passer le test du « *cool moment* »[31].

---

[30] La peur est associée à l'activation de noyaux dans l'amygdale, une zone très ancienne du cerveau. LeDoux a montré que l'amygdale est capable d'associer la peur à des types de stimuli obtenus par perception tactile, olfactive, visuelle, auditive , sans utiliser des fonctions du cortex cérébral, donc sans amorcer un processus de pensée consciente (LeDoux : 1998).

[31] Une des cibles principales de Gert est la conception que Rawls se fait de la planification rationnelle de sa vie. Pour faire court, Rawls soutient qu'un plan de vie rationnel est choisi en fonction des principes abstraits du choix rationnel – cohérence, efficience, inclusivité, maximisation, etc. – et dans un contexte de délibération complète (*full deliberative rationality*). La délibération complète est censée produire un jugement éclairé par l'ensemble des considérations

Je mets de côté ici la question de savoir si la critique de Gert est correcte. Il existe des données empiriques beaucoup plus solides à l'encontre du modèle de rationalité vulcain et de sa version « *cool moment theory* ». Non seulement il appert qu'on ne prend pas ou très peu de décision importante dans un état de « froideur » totale, mais il semble que les agents qui se trouvent – à la suite de la formation de lésions dans certaines zones de leur cerveau – dans un perpétuel « *cool moment* » souffrent d'un déficit important de contrôle de soi qui les poussent à adopter des comportements impatients, voire impulsifs.

Le neurologue Antonio Damasio est célèbre pour avoir mis en relief les déficits émotionnels et leurs conséquences chez des sujets ayant subi des lésions cérébrales. En étudiant le comportement et les capacités cognitives de patients qui avaient subi des lésions importantes dans les lobes préfrontaux et dans certaines couches sous-corticales, Damasio a pu mesurer le rôle joué par les émotions dans la prise de décision (Damasio, 1995, 2010). Il est établi depuis les premières cartographies du cerveau que les aires responsables du contrôle de soi se situent dans le cortex frontal. Damasio a établi que, bien que les zones plus anciennes, que sont les aires sous-corticales, aient un rôle important à jouer dans la production et la perception des émotions, les zones plus modernes du cerveau y jouent également un rôle. Contrairement à l'opinion commune, il n'y a pas de traitement de l'information en vue d'une décision dans les aires modernes comme le néocortex et d'ajout subséquent d'une couleur émotive. Fustigeant cette opinion, Damasio la décrit de la manière suivante :

> « *Nous nous représentons généralement la capacité de ressentir des émotions comme une faculté mentale surnuméraire, un à-côté de la pensée rationnelle, voulu par la nature, mais*

---

pertinentes. Ce jugement identifie un plan de vie qui inclut nécessairement, selon Rawls (1973 : 457-464), un désir pour les biens premiers. Les biens premiers – sécurité, opportunité, santé, intelligence, vigueur, imagination, etc. – sont des choses que tous les agents rationnels doivent désirer parce que ces biens offrent des conditions favorables à la poursuite de leurs objectifs et cela peu importe en quoi ces derniers consistent. Gert soutient que Rawls, avec sa condition de délibération complète, adhère à la théorie du « *cool moment* ». Il avance que la théorie du « *cool moment* » ne permet pas de justifier rationnellement une préférence pour les biens premiers et une aversion pour le manque de ces biens. Un agent peut, dans un état d'esprit calme et serein, analyser les conséquences d'une ablation volontaire d'un membre pourtant en santé et procéder ensuite, dans un état calme ou émotif, à l'ablation du membre en question. On jugera intuitivement le désir de cet agent irrationnel, pourtant il passe le test du « *cool moment* ».Il s'agit peut-être d'un cas limite. Mais on peut trouver beaucoup d'exemples communs de décision motivée par des désirs qui ne passent pas le test, mais qui sont intuitivement rationnels. Ce qui montre que non seulement la théorie du « *cool moment* » n'est pas assez restrictive pour certaines catégories de désirs, mais qu'elle est trop restrictive pour d'autres. Par exemple, beaucoup de femmes enceintes souhaitent éviter l'épidurale afin d'obtenir un accouchement le plus « naturel » possible. Seulement, la plupart d'entre elles se rebiffent lorsque la douleur des contractions devient aiguë. Certaines d'entre elles le regrettent même après, lorsqu'elles réintègrent leur état plus froid. Doit-on considérer leur désir d'obtenir une injection contre la douleur comme irrationnel parce qu'il ne passe pas le test du « *cool moment* »? « Gardez la tête froide! » relève plutôt d'une injonction utile dans nombre de situations qu'une propriété constitutive du comportement rationnel.

> *non pas le sujet pensant. S'il s'agit d'émotions agréables, nous les ressentons comme un luxe; si elles sont désagréables, nous les endurons comme des perturbations non souhaitées.* » (1995 : 79)

Damasio en est venu à la conclusion que les émotions constituent des éléments essentiels à la prise de décisions dans un nombre considérable de situations en étudiant le cas d'Elliot. Elliot est un patient dont une tumeur cancéreuse a comprimé et détruit une partie du cerveau, mais fut extraite avec succès. Les proches d'Elliot observèrent alors un changement radical de personnalité. Auparavant, homme d'affaires compétent et consciencieux, Elliot se mit à prendre des risques inconsidérés en s'associant avec des personnes « douteuses » pour mener des affaires également « douteuses ». Elliot perdit son emploi et éprouva des difficultés importantes à diriger sa vie. Il semblait même éprouver des difficultés importantes à prendre des décisions aussi anodines que de choisir une place où garer sa voiture ou choisir un restaurant. En fait, chaque fois qu'Elliot devait faire face à une situation problématique dans le cadre de ses activités quotidiennes, il demeurait étrangement longtemps dans un état d'indécision. Et lorsqu'il se décidait enfin, c'était souvent à la faveur d'une ligne de conduite inappropriée pour laquelle il manifestait une obstination incompréhensible. Au travail, il pouvait passer une journée entière à classer les documents sur son bureau. Damasio lui fit passer toute une batterie de tests d'intelligence visant à mesurer toutes sortes de capacités cognitives comme le maintien de l'attention, la mémoire de travail, la logique et la capacité de résoudre des problèmes inférentiels à partir d'informations limitées, etc. La surprise de Damasio fut grande lorsqu'il se rendit compte que son patient obtenait des résultats soit moyens soit au-dessus de la moyenne pour l'ensemble de ces tests. Elliot ne souffrait d'aucun déficit au niveau de ces compétences cognitives les plus froides. En revanche, Elliot exhibait constamment une humeur calme et ne semblait pas être particulièrement affecté par sa condition. En fait, Elliot ne semblait pas éprouver la moindre émotion. Il parlait de lui-même et de ce qui lui arrivait comme un spectateur impartial. Après l'étude de nombreux cas analogues, à l'aide notamment d'appareils de résonnance magnétique, Damasio en arrive à la conclusion que les émotions jouent un rôle essentiel dans la résolution de problèmes décisionnels concernant la manière dont un agent peut améliorer ou éviter d'altérer ses conditions d'existence, en particulier son statut social. Lorsque les problèmes à résoudre nécessitent un degré de détachement important – comme la résolution de problèmes de logique formelle ou de statistique – les agents dont les circuits émotionnels sont altérés obtiennent des résultats comparables à ceux

dont les mêmes circuits sont intacts. Les choses se corsent lorsque les problèmes à résoudre concernent le bien-être de ces personnes. Il semble que dans ce domaine de compétence, ce que Damasio appelle les « marqueurs somatiques », permettent d'orienter et d'encadrer efficacement le processus de délibération.

L'hypothèse des marqueurs somatiques provient des résultats obtenus par des sujets normaux et par des sujets atteints de lésions à un test de sélection de cartes appelé le jeu de l'Iowa – qui est un jeu de hasard. Ce test a été développé pour mesurer et évaluer les déficits des patients atteints de lésions en les plaçant dans un contexte décisionnel comportant de l'incertitude. Dans ce test, les sujets doivent choisir une carte dans un des 4 paquets disposés devant eux. Ils doivent effectuer au total 100 choix. On les questionnait après les 20 premiers coups sur leur compréhension du jeu, puis chaque 10 coups. À chaque fois que le sujet tourne une carte, il obtient ou perd le montant indiqué. Deux paquets impliquent statistiquement plus de pertes que de gains, tandis que les deux autres offraient statistiquement plus de gains que de pertes. La distribution des gains et des pertes dans les séries est telle que les mauvais paquets présentent plus de gains en début qu'en fin de série, alors que c'est l'inverse pour les bons paquets. Pour chacun des coups, on mesure la conductance de la peau pour savoir si le sujet a une réaction affective. Les résultats sont importants. Les sujets avec lésions ont tendance à choisir les mauvais paquets, en dépit du fait qu'ils arrivent à comprendre le jeu et qu'ils savent qu'il y a des mauvais paquets, cette connaissance ne les motive pas à choisir les bons paquets. Les mesures de conductibilité de la peau de ces sujets montrent qu'ils n'éprouvent pas d'émotion pendant le processus de délibération. En revanche, on observe une réaction émotionnelle chez les sujets normaux *avant* même qu'ils arrivent à comprendre le jeu. Ces sujets commencent dès lors à choisir les paquets plus avantageux et n'« attendent » pas de comprendre la structure des gains et des pertes, et leur répartition statistique au travers des paquets.

Cette expérience montre que les agents qui doivent faire des choix en situation de risque et pour lesquels les gains et les pertes importent (ils ne sont pas dans une situation de choix impartial), adoptent des lignes de conduite rationnelle que si et seulement si les bons choix et les mauvais choix sont émotionnellement marqués. Le défaut de réaction émotionnelle

n'entraîne pas un contrôle de soi plus accru, c'est plutôt l'inverse qui se produit[32]. L'autre conclusion à tirer est que les états émotionnels qui motivent certains de nos choix peuvent aisément passer sous le radar de la conscience. Cela ne jette pas seulement un certain discrédit sur nos mécanismes d'introspection, cela relativise aussi l'impact des raisonnements conscients dans la prise de décisions rationnelles.

Cela dit, ces conclusions n'impliquent pas en retour que les agents akratiques soient nécessairement en déficit émotionnel. Seulement, s'il y a un déficit émotionnel, il y a une probabilité accrue d'épisodes de faiblesse de volonté. Cela n'implique pas non plus que les émotions ne jouent qu'un rôle positif dans le processus décisionnel des agents. Il ne fait pas de doute que le type d'émotion et son intensité jouent un rôle dans certains comportements akratiques. Il n'est pas exclu qu'on explique certains épisodes de faiblesse de volonté par l'emprise d'une émotion forte et de la mauvaise qualité du processus délibératif qu'elle contamine – un peu comme le voit Aristote.

D'ailleurs, on sous-estime, semble-t-il, même la profondeur du changement que les émotions sont susceptibles d'entraîner dans nos propres dispositions à adopter des comportements allant à l'encontre de nos propres intérêts et de nos principes moraux les plus forts. L'économiste Dan Ariely et le psychologue George Loewenstein ont mesuré la profondeur de ce changement à l'aide d'une expérience menée sur des étudiants universitaires mâles du MIT (2006 : 87-98). On demandait aux cobayes de mesurer subjectivement leur disposition à accomplir des actes à caractère sexuel. Ces actes appartenaient à différentes catégories, allant de la mesure du degré d'attraction d'activités comme avoir des rapports sexuels avec quelqu'un d'extrêmement gros ou envers lequel on entretient une haine viscérale à la probabilité d'accomplir des actes moralement répréhensibles, comme encourager une femme à boire de l'alcool ou la droguer pour maximiser ses chances d'avoir des rapports sexuels avec elle. Le premier groupe de cobayes devait effectuer cette mesure d'abord « à froid », c'est-à-dire dans un état affectif calme. Le second groupe devait effectuer la mesure dans un état d'auto-excitation sexuel avancé. Les résultats sont surprenants. Nous savons intuitivement que nous sommes tous

[32] En fait, comme Damasio a tenu à préciser, les dommages aux régions corticales mentionnés plus haut n'abolissent pas vraiment les réactions émotionnelles, pas plus que l'activation de ces régions suffit à produire une réponse émotionnelle forte. Plutôt, ces dommages altèrent la connexion cognitive entre les réactions émotives et l'anticipation d'événements et les conséquences des décisions et des stratégies de choix basés sur des émotions (Bechara, Damasio, & Damasio, 2000).

davantage disposés à accomplir certains actes lorsque nous nous trouvons dans un état d'excitation sexuel, mais la différence est très marquée dans certains cas (ex. : on observe une augmentation de 125% de la disposition (*willingness*) à tenter d'obtenir des rapports sexuels avec une femme qui a pourtant clairement signifier un refus). Ce qui fait dire à Ariely qu'en chacun de nous vivent à la fois un docteur Jekyll et un Mr. Hide (2008 : 87-108). Seulement, notre situation « à froid » est peut-être pire que celle du pauvre docteur, puisque nous éprouvons – et c'est un des résultats les plus intéressants de l'expérience – de la difficulté à prédire à partir de notre état d'esprit actuel, les comportements que nous serons susceptibles d'adopter lorsque celui-ci se modifiera.

2.2 La faiblesse de volonté dans la tradition médiévale chrétienne

Le traitement que la faiblesse de volonté a reçu dans la littérature médiévale diffère de ce qu'on trouve chez Platon et Aristote. Des auteurs comme Augustin, Thomas d'Aquin, Anselme et Burridan, ne détachaient pas clairement les questions théoriques portant sur la motivation des questions portant sur la morale – et la religion. Dans l'optique de Platon et d'Aristote, la rationalité est un trait constitutif de l'esprit humain (elle relève de *ce qu'on fait*), tandis que les erreurs et les égarements seraient des accidents (ils relèvent de *ce qui nous arrive*). Dans la tradition médiévale, les choses sont inversées. Celle-ci voit dans la constitution de notre esprit source de nos égarements, classés sous la rubrique de la faiblesse de la « chair », alors que les sources de rationalité sont attribuables au Divin (Ainsli, 2001). L'existence du péché originel, l'explication du mal dans le monde et la doctrine du libre arbitre, ont mené les auteurs de l'époque à inférer l'existence d'une faculté nouvelle, la *volonté* (Pironet & Tappolet, 2003). Dans ce qu'on a par la suite appelé le *volontarisme* médiéval, la volonté est dès lors conçue comme la faculté qui est capable d'exercer le libre arbitre nécessaire pour choisir entre le bien et le mal. Aussi, c'est dans la volonté que réside la possibilité du salut de son âme autant que la possibilité de sa déchéance. D'une manière plus prosaïque, le rôle que joue la volonté dans le volontarisme est celui d'ultime décideur, le « tiers » acteur qui tranche entre les désirs issus de la passion et les exigences de la raison. Il revient en quelque sorte aux passions et à la raison de faire des propositions, et à la volonté d'en disposer.

2.2.1 Le volontarisme : une conception naïve?

Bien que le volontarisme motivationnel apparaisse d'abord et avant tout comme une doctrine d'inspiration théologique, il n'est pas dénué d'intérêt psychologique. Il met l'emphase sur la nature conflictuelle de la stricte akrasie et s'accorde même avec nos intuitions introspectives selon lesquelles nous sentons, d'une certaine façon, que nous choisissons librement entre des considérations incompatibles lorsque nous adoptons des comportements akratiques ou lorsque nous y résistons. Il a aussi un intérêt philosophique dans la mesure où il propose une alternative claire au cognitivisme motivationnel. Les agents peuvent agir à l'encontre de ce qu'ils estiment le plus approprié de faire parce qu'ils sont tout simplement libres de le faire.

On a tenté de voir dans le volontarisme une conception qui soulève plus de problèmes théoriques qu'elle n'en résout, et, de surcroît, il s'agit de problèmes insolubles. Elle semble avoir peu de puissance explicative parce qu'elle pose l'hypothèse d'une faculté qui choisit entre des considérations sans être elle-même déterminée par des considérations. On postule une autodétermination absolue de la volonté, ce qui a pour résultat que ce qu'on souhaite au départ expliquer, l'akrasie ou la stricte akrasie, n'est pas du tout expliqué parce qu'on ne sait pas pourquoi la volonté choisirait tantôt des lignes de conduite appropriées, tantôt des lignes de conduite inappropriées. Si l'on souhaite expliquer les choix de la volonté, alors on s'expose à une régression à l'infini (Ryle, 1949). On devra – conformément à l'esprit du volontarisme – postuler une métavolonté supplémentaire, répéter ensuite la procédure conceptuelle pour expliquer les choix de la métavolonté, et ainsi de suite.

Cela dit, le concept de volonté n'est évidemment pas à jeter aux orties. Il a reçu dans l'histoire de la philosophie un traitement varié qui recoupe toutes sortes de problématiques allant de la dynamique des impulsions motrices issues d'un système centralisé, à la responsabilité morale en passant par l'idéal d'autonomie, du contrôle de soi et de la nature de l'expérience subjective du moi[33]. Aussi, il n'est pas toujours facile de démêler les usages

[33] C'est ce qui a fait dire à Nietzsche : « *« Vouloir » me semble être, avant tout, quelque chose de compliqué, quelque chose qui ne possède d'unité qu'en tant que mot* […]. […] *dans tout vouloir il y a, avant tout, une multiplicité de sensations qu'il faut décomposer : la sensation du point de départ de la volonté, la sensation de l'aboutissant, la sensation du « va-et-vient » entre ces deux états ; et ensuite une sensation musculaire concomitante qui, sans que nous mettions en mouvement « bras et jambes », entre en jeu dès que nous « voulons ». De même donc que des sensations de diverses sortes sont reconnaissables, comme ingrédients dans la volonté, de même il y entre, en deuxième lieu, un ingrédient*

du terme. Mais il est difficile de voir si le concept de volonté des auteurs médiévaux correspond davantage aux notions plus modernes de désir, ou bien d'intention ou encore de décision (ou de choix). Par contre, le postulat d'autodétermination absolue est clairement erroné, peu importe l'interprétation qu'on veut bien choisir.

2.2.2 Faiblesse de volonté ou faiblesse motrice?

La tradition médiévale a orienté son attention davantage sur l'aspect exécutoire des volitions que sur leur formation. Aussi, les problèmes relatifs au raisonnement pratique, à la délibération et au jugement jouent un rôle secondaire par rapport à la manière dont la volonté, une fois formée, cause l'action ou les mouvements mécaniques du corps[34]. Selon Pironet et Tappolet, le concept de faiblesse de volonté décrit mieux la difficulté pratique que vit l'agent lorsqu'il échoue à mobiliser les ressources internes nécessaires pour *exécuter une action voulue*, tandis que le concept d'akrasie décrirait mieux les difficultés à *former le vouloir approprié au jugement* de l'agent (Pironet & Tappolet, 2003). C'est ce qui distingue l'approche médiévale de l'approche de Platon et d'Aristote.

La distinction est éclairante, mais à condition qu'on interprète cependant le « *vouloir* » de la volonté comme une sorte de décision, de choix ou d'intention de faire quelque chose, et non pas comme une sorte de désir. Si vouloir est une décision de la volonté, alors une faiblesse de volonté relèverait d'une difficulté à exécuter au niveau moteur une décision. Il s'agirait donc d'un problème neurologique lié à la formation de commandes dans le cerveau [35] destinées aux structures musculo-squelettiques ou à l'exécution de ces commandes dans ces structures, comme c'est le cas par exemple pour des problèmes de bégaiement et pour la maladie de Parkinson. Dans ce cas, l'akrasie et la faiblesse de volonté seraient des phénomènes radicalement différents.

Mais si l'on interprète le vouloir comme une sorte de désir – comme le propose une analyse de Davidson – alors, la faiblesse de volonté ne relève pas d'un problème moteur, mais plutôt d'un problème de motivation (1978 : 136-145). Dans ce cas, l'akrasie et la faiblesse

---

nouveau, la réflexion. Dans chaque acte de la volonté il y a une pensée directrice. Et il faut bien se garder de croire que l'on peut séparer cette pensée du « vouloir », comme s'il restait encore, après cela, de la volonté ! » (1886 : § 19).

[34] Dans l'Épître aux Romain 7, Paul témoigne : *«Je vois une autre loi dans mes membres qui combat contre les lois de mon esprit.»*.

[35] En particulier dans les noyaux gris centraux et dans le thalamus. À cet égard, les noyaux gris jouent un double rôle. Ils fournissent les impulsions nécessaires aux mouvements, mais sont capables de les inhiber en période de repos.

de volonté seraient des phénomènes qui se recoupent – partiellement ou en quasi-totalité – mais pourraient se situer à des niveaux différents de l'échafaudage motivationnel et la distinction de Pironet et Tappolet devient plus superficielle.

Il y a cependant de bonnes raisons de croire que c'est la première interprétation qui est correcte. Le célèbre argument de l'âne de Buridan repose sur une interprétation du vouloir de la volonté comme une décision et non comme un désir. Un âne assoiffé, situé à égale distance entre deux sources d'eau également attrayantes, ne mourra pas de soif, mais optera pour une des deux alternatives (Saarinen : 1994). Il s'agit initialement d'un argument en faveur de l'hypothèse de l'autodétermination absolue de la volonté, mais il exprime bien l'idée que le vouloir de la volonté est une sorte de décision et non pas une sorte de désir, puisque ce qui tranche en faveur d'un choix ici ne relève pas des désirs (ou même de la cognition).

Distinguer les deux sens du concept de vouloir, nous permet d'éviter beaucoup de confusion. Si vouloir quelque chose se situe en amont d'une décision, d'un choix effectif ou d'une intention, alors le phénomène de la faiblesse de volonté relève de problèmes motivationnels au sein de l'agent et pose, par conséquent, une énigme motivationnelle. En revanche, si vouloir est quelque chose qui se situe en aval d'une décision, d'un choix effectif ou d'une intention, alors la faiblesse de volonté, envisagée comme un problème de connexion entre la décision, le choix ou l'intention, avec les systèmes moteurs, ne relève pas de la motivation comme telle.

Cela dit, je pense que lorsqu'on parle de manière non analytique de faiblesse de volonté, on traite le vouloir comme un élément motivationnel. La motivation concerne les raisons, au sens large (incitations, procédure décisionnelle, préférences, stratégie de contrôle de soi, émotions, croyances, stimuli, aversions, etc.) de décider ou de former l'intention de faire quelque chose. Aussi, une investigation des facteurs motivationnels conduisant les agents à prendre des décisions contre-productives – ou allant carrément à l'encontre de leur intérêt bien pesé – comme le sont les démarches de Platon et d'Aristote, ont plus de pertinence pour cerner le problème de la faiblesse de volonté (au sens non strict du terme) que les considérations sur les problèmes de transmission motrice. Trop manger, fumer, procrastiner, surconsommer, s'endetter à l'excès, ne pas suffisamment se protéger lors de rapports sexuels, remiser nos appareils d'entraînement physique, ne pas faire des

économies, sont souvent des problèmes de faiblesse de volonté, mais rarement des problèmes d'impulsion motrice.

2.3 Conclusion

Bien qu'erroné, le cognitivisme motivationnel de Platon est une position théorique beaucoup plus robuste et subtile qu'il y parait. Nous surestimons spontanément notre capacité à évaluer correctement l'intensité et la durée de nos états hédoniques futurs, et il n'est pas exclu que cette myopie cognitive joue un rôle important dans l'explication de l'akrasie. Le fait que la plupart de nos désirs et un grand nombre de nos émotions aient des antécédents cognitifs rend également plausible le cognitivisme motivationnel. Toutefois, il est difficile de défendre l'hypothèse que le soubassement motivationnel d'un agent est constitué exclusivement de considérations cognitives, et que, par conséquent, il agit toujours par jugement et non pas désirs ou préférences. D'ailleurs, le rejet du cognitivisme motivationnel ouvre la porte à la stricte akrasie. Par contre, l'hypothèse de la maximisation comme norme constitutive de la rationalité pratique continue à être un élément de tension important pour toute théorie qui accepte l'existence de la stricte akrasie.

Les modèles explicatifs que propose Aristote sont plus variés, mais pas nécessairement incompatibles avec celui de Platon. La négligence et le raisonnement incomplet peuvent expliquer dans certains cas pourquoi les agents échouent à améliorer leur sort, tandis la contamination affective du raisonnement pratique serait un meilleur candidat pour expliquer les cas plus particuliers de stricte akrasie. Or, contrairement à ce qu'Aristote pensait, il s'avère que les émotions sont en fait nécessaires pour prendre des décisions rationnelles – du moins des décisions qui concernent directement l'agent. Par conséquent, l'hypothèse d'une contamination affective du raisonnement pratique doit être rejetée, ou du moins révisée.

Abstraction faite du caractère suranné des raisons théologiques qui justifient le volontarisme médiéval, ce dernier offre tout de même une réponse stérile au problème de la faiblesse de volonté en général et de la stricte akrasie en particulier. Postuler l'existence d'une faculté de choix complètement autodéterminée implique certes la possibilité de l'existence de la stricte akrasie, mais rend tous les choix inexplicables par des antécédents motivationnels.

L'existence de la stricte akrasie est devenue le principal défi théorique que pose le problème de la faiblesse de volonté depuis les premières tentatives d'explication dans l'Antiquité et au Moyen-âge. Depuis le renouveau des enquêtes philosophiques sur le sujet initié au milieu du 20^{e} siècle, les philosophes ont fait porter principalement leurs efforts théoriques dans la solution ou la dissolution du problème de la stricte akrasie. C'est pourquoi j'aborderai dans le prochain chapitre ce qui est devenu la conception Standard de la stricte akrasie, et qui sert maintenant de point de référence pour les débats actuels.

3

La conception Standard : jugement, préférence et incohérence

Ce que l'incontinence a de particulier est que l'agent ne parvient pas à se comprendre lui-même; il reconnaît, dans son comportement intentionnel, quelque chose d'essentiellement sourd.

–Donald Davidson, Actions et événements

3.1 La conception Standard de la stricte akrasie

À l'époque de Platon et d'Aristote, on concevait plutôt confusément le phénomène de la faiblesse de volonté. On avait l'intuition forte qu'il s'agissait d'une faute à l'égard des canons de la raison et non strictement un égarement moral – et il s'agissait là d'une avancée considérable dans l'identification du phénomène. Seulement, on avait tendance à confondre le problème avec d'autres types de comportements plus ou moins rationnels. Comme je l'ai mentionné dans le chapitre précédant, Aristote expliquait, par exemple, certains comportements « akratiques » par un manque de connaissance à l'égard des faits pertinents à la situation. Aussi, les Anciens avaient une conception à la fois vague et étriquée de la rationalité humaine, incompatible avec les émotions. Les émotions et autres passions de l'âme, étaient vues comme quelque chose qui arrive *de l'extérieur* aux personnes, par ailleurs conçues comme rationnelles par essence. Ils avaient donc tendance à pointer les émotions non seulement pour expliquer les cas putatifs d'akrasie mais également pour les diagnostiquer.

L'état de la recherche philosophique sur le sujet n'a malheureusement pas fait de progrès significatifs pendant plusieurs siècles. Les auteurs chrétiens ont placé le problème sous la rubrique de la « *faiblesse de la chair* » et, à la différence des Anciens, ont plutôt vu l'akrasie comme une propriété constitutive des personnes et la raison comme une propriété constitutive de Dieu. En dépit de certaines tentatives moyenâgeuses et modernes de définir clairement le phénomène, il a fallu attendre jusqu'au 20^{e} siècle pour voir se dessiner un portrait beaucoup plus précis du phénomène. La multiplication des travaux, des articles et des ouvrages sur le sujet, encouragée par l'émergence des théories philosophiques de

l'action, a favorisé une convergence dans la définition de l'akrasie et des problèmes théoriques qu'elle suscite. Aussi, le théoricien se retrouve maintenant avec une conception Standard à partir de laquelle il travaille.

Il y a plusieurs variantes de la conception Standard, mais je pense que celle que Davidson examine dans l'article intitulé « *Comment la faiblesse de volonté est-elle possible?* » en constitue sans doute la version canonique :

Conception Standard :

(CS) : *En faisant* x*, un agent agit de manière incontinente, si et seulement si : a) l'agent fait* x *intentionnellement ; b) l'agent croit qu'il* y *a une autre action possible* y *à sa portée ; et c) l'agent juge que, tout bien considéré, il serait mieux de faire* y *que de faire* x*. (1970 : 38)*

Cette définition est censée être construite de manière suffisamment générale et moralement « neutre » pour englober tous les cas putatifs de faiblesse de volonté et exclure les cas qui n'en relèvent intuitivement pas. La clause (a) exclue les actions accomplies sans que l'agent ait conscience de certains de ses aspects pertinents[36]. Certains des cas qu'Aristote traite comme des instances d'akrasie (ex. : Un alcoolique en période de sevrage qui boit le contenu d'une tasse remplie de Margarita en pensant qu'il s'agit de jus de fruit.) ne satisfont pas la définition. La clause (b) permet d'exclure également des cas intuitivement non-akratiques (ex. : Un ex-alcoolique qui boit intentionnellement un punch alcoolisé parce qu'il n'y a pas autre chose à boire et qu'il souffre de déshydratation). Tandis que la clause (c) stipule que l'agent doit avoir formé un jugement selon lequel une ligne de conduite qu'il

[36] En fait, Davidson traite l'expression « *faire x intentionnellement* » comme une description *sous laquelle* une action est intentionnelle pour l'agent qui l'accomplit, et non comme une description qui réfère à une action et identifie une de ses propriétés circonstancielles (Davidson, 1971). Pour Davidson, et la plupart des théoriciens de l'action, toutes les actions sont intentionnelles, sans quoi il ne s'agit pas d'action, mais de « simples » comportements. Seulement, les actions n'apparaissent pas intentionnelles sous toutes les descriptions. Certaines descriptions incluent des indications sur des conséquences de l'action alors que d'autres non. Par exemple, affirmer que Fulgence à tuer intentionnellement Adélard n'est pas la même chose que de dire que Fulgence a bougé le doigt ou tiré sur la gâchette, bien que ces trois descriptions portent sur une seule et même action : un mouvement corporel de Fulgence causé par ses croyances et ses désirs. Le fait d'inclure dans nos descriptions d'action des indications sur certaines de leurs conséquences a été initialement désigné par Joel Feinberg (1970 : 119-151) comme *l'effet accordéon*. On peut parler d'effet accordéon lorsqu'on décrit une action en indiquant des conséquences de celle-ci que l'agent lui-même anticipait, croyait causer, visait par son action. Si Fulgence a intentionnellement tué Adélard, alors Fulgence estimait qu'il causerait sa mort en appuyant sur la gâchette.

Beaucoup de théoriciens rejettent cette analyse. Les partisans de la description « à grains fins » de l'action, comme Goldman (1970) et Kim (1976), défendent l'idée que des descriptions comme celles mentionnées portent sur des actions différentes (conçues comme des exemplifications de propriétés) et non sur une même action décrite de manière différente. Ces théoriciens prennent donc à la lettre l'opinion commune qu'il y a des actions intentionnelles et non intentionnelles.

n'adopte pas est meilleure, plus avantageuse, plus judicieuse, plus adéquate, etc. que celle qu'il adopte. Cela nous permet d'exclure des cas où l'agent s'engage dans une ligne de conduite clairement sous-optimale parce qu'il n'a pas été capable ou en mesure d'en identifier une meilleure.

Aussi, la nature du jugement auquel on fait référence dans la définition standard est importante. Il n'est pas question ici de jugement de fait dans le sens ordinaire du terme, mais de jugements sur ce que l'agent estime devoir faire. Si un agent juge qu'il est meilleur, plus avantageux, plus adéquat, etc. de faire *x*, alors il est préférable selon lui de faire *x*. Cela n'implique évidemment pas que l'agent ait une sagesse totale concernant ce qui est préférable de faire dans l'absolu – auquel cas on agirait la plupart du temps de manière akratique. Mais, *compte tenu* des informations dont il dispose et de ses aspirations, désirs et aversions, l'agent préfèrera une ligne de conduite à toute autre – ou du moins, sera indifférent entre plusieurs lignes de conduite (ex æquo) qui sont cependant préférables à toutes les autres.

Une version simplifiée de la définition standard de la faiblesse de volonté consiste à dire qu'un agent est akratique si et seulement s'il agit à l'encontre de son meilleur jugement. Ici, le meilleur jugement de l'agent correspond non pas à un jugement qu'il estime être meilleur, mieux informé, mieux appuyé que les autres jugements, mais un jugement à l'effet qu'une ligne de conduite est préférable à toutes les autres. Ce point semble toutefois constituer un écart avec la définition standard longue. Mais il n'en est rien. Si un agent estime que *x* est préférable à *y*, cela n'implique pas que pour lui *x* est préférable à *z* lorsque *z* fait partie des options en lice. Il se peut que *z* soit l'option préférable à toutes autres. Cependant, si l'agent qui considère ces options n'opte pas pour *z* – ce que lui indique son meilleur jugement – alors il sera réputé akratique parce qu'il satisfait la définition standard longue. En d'autres termes, la définition standard longue implique qu'un agent akratique agisse à l'encontre de son meilleur jugement et que tout agent qui agit à l'encontre de son meilleur jugement est akratique. Aussi, la version simplifiée est différente de la version longue seulement en ce qu'elle ne mentionne pas explicitement les clauses (a) et (b). J'utiliserais donc indistinctement les deux versions pour le reste de mon propos.

3.1.1 Précisions concernant la conception standard

Même si la définition standard est claire, certaines précisions sont néanmoins nécessaires. On ne peut pas dire que l'agent agit intentionnellement à l'encontre de son meilleur jugement si son comportement est simplement *incohérent* (*inconsistant*) avec son meilleur jugement (Audi, 1993).

On ne peut pas dire, par exemple, qu'un agent vit un épisode akratique s'il n'applique pas le résultat de sa délibération parce qu'il l'a oublié. On peut délibérer pour résoudre un problème décisionnel futur et obtenir un résultat qu'on oubliera ensuite dans le feu de l'action, ou dont on oubliera certains des détails. Je peux, par exemple, délibérer pour déterminer quel chemin je dois prendre pour me rendre chez un ami. Comme l'itinéraire choisi est passablement sinueux et que je ne dispose pas de GPS ou de carnet de note pour l'inscrire, j'en oublierai certaines portions et m'engagerai dans des avenues inappropriées. Il est également possible qu'ayant délibéré sur la manière dont je dois aborder quelqu'un de particulièrement susceptible, j'oublie en cours de route la ligne de conduite adéquate et réintègre progressivement mes habitudes conversationnelles nonchalantes.

Je peux également garder à l'esprit le résultat de ma délibération antérieure, mais, sans en oublier le détail, accomplir un acte qui n'exemplifie pas les propriétés souhaitées par simple ignorance, distraction, légèreté, étourderie, inadvertance, mégarde, imprudence ou négligence. Je peux savoir que je suis en mesure d'accomplir une action qui exemplifie des propriétés souhaitées – et ce faisant respecter les contraintes endogènes et exogènes inhérentes à mes choix –, mais ne pas la piloter correctement. Il s'agit d'un problème d'exécution en temps réel d'un scénario choisi par ignorance et non pas un problème de faiblesse de volonté.

Je sais que ce dernier point peut être critiqué. Les personnes qui souffrent de faiblesse de volonté de manière aiguë – les procrastinateurs invétérés et les grands impatients – sont souvent, comme la recherche en psychologie clinique l'a montré, facilement distraites et font souvent preuve de négligence dans leur attitude et leur comportement. La capacité de maintenir son attention sur une tâche cognitive ou une activité complexe de manière suffisamment soutenue est une condition nécessaire pour en assurer le succès. Mais on peut traiter, à mon avis, la portée de cette capacité comme une contrainte endogène au choix des

scénarios. Si un agent choisit un scénario qu'il est incapable, *considérant sa capacité d'attention*, de réaliser, alors il ne se montrera pas akratique s'il ne se conforme pas à son meilleur jugement (voir la notion de contrainte endogène dans le prochain chapitre). On jugera que sa planification était, pour lui, trop ambitieuse. Cela a sans doute pour effet de restreindre de façon significative la quantité de cas putatifs de faiblesse de volonté[37]. Peut-être un peu trop. Mais il est clair que la présence du facteur « ignorance » peut nous servir à rejeter un diagnostic de faiblesse de volonté. Dans quelle mesure? Cela est une question qui, à ce que je sache, n'a pas été suffisamment abordée, même si elle est importante.

Cela dit, ce n'est pas parce que l'agent dispose d'une solution à un problème qu'il doit l'appliquer en dépit du fait qu'il peut acquérir entre temps de nouvelles informations pertinentes. Je peux, par exemple, choisir un itinéraire de voyage qui implique de visiter beaucoup de lieux qui offrent des activités de plein air. Mais si j'apprends, avant de partir, que les prévisions météorologiques sont plutôt défavorables, je serai enclin à suspendre ma décision ou à réviser mon plan pour les vacances. Dans ce cas, il se peut que mon meilleur jugement ait changé : je juge maintenant qu'il est préférable pour moi de ne pas suivre cet itinéraire. Aussi, je continue à me conformer à mon meilleur jugement *le plus récent*. Or, il se peut que je n'aie pas encore rejeté mon itinéraire, étant donné l'incertitude des prévisions météo pour les jours à venir, mais que je sois tout simplement en train de reconsidérer ma décision. Il est toujours préférable pour moi que je me conforme à l'itinéraire prévu, en dépit des prévisions météo, mais je ne décide pas tout de suite de m'y engager. Suspendre l'accomplissement d'une action pourtant conforme à un meilleur jugement ne relève pas toujours d'un épisode de faiblesse de volonté. Nous pouvons avoir de bonnes raisons de le faire, même si nous ne disposons pas nécessairement d'une solution de rechange.

La conception Standard porte *exclusivement* sur la forme strictement akratique de la faiblesse de volonté et non sur sa forme diachronique. De manière générale, la faiblesse de volonté est un phénomène qui consiste à la dégradation volontaire de nos propres conditions de vie – ou du moins au refus de ne pas les améliorer, alors qu'il serait facile de le faire. On parlera alors de faiblesse de volonté au sens large et non technique du terme. La

[37] On peut, je pense, se servir de la suggestion qu'Audi a faite à l'égard des comportements compulsifs et classer les cas où l'agent ne s'est pas conformé à son meilleur jugement parce qu'il est incapable de maintenir son attention de manière suffisamment soutenue sur une tâche le requérant dans la catégorie de cas de faiblesse *dans* la volonté.

stricte akrasie serait un phénomène plus restreint d'incohérence entre un jugement et une action.

Bien que la stricte akrasie donne régulièrement lieu à une dégradation des conditions de vie de l'agent, elle n'engendre pas *nécessairement* cet effet. Un agent strictement akratique peut agir en conformité avec son intérêt, mais en faisant valoir celui-ci à l'encontre de l'intérêt d'autrui dans des cas de stricte akrasie morale. Quand les agents agissent à l'encontre de principes moraux qu'ils estiment supérieurs à leur propre intérêt du moment, ils vivraient théoriquement un problème de stricte akrasie.

3.1.2 Deux révisions mineures et nécessité d'une conception canonique plus générale

On a de bonnes raisons de penser que la conception Standard de l'akrasie (CS), qui contraste le meilleur jugement de l'agent et son action, est incorrecte. Elle est un peu trop étriquée et est en plus un facteur de confusion. C'est la notion d'« action » qui est responsable de ces défauts.

Par contraste, la notion de « décision » est plus générale et aussi plus appropriée pour décrire le phénomène. On ne doit pas considérer seulement les cas où l'agent *agit* à l'encontre de son meilleur jugement, mais ceux pour lesquels il *décide* à l'encontre de ce dernier. Je peux former le jugement selon lequel qu'il est préférable, toutes choses considérées, que je limite ma consommation d'aliments sucrés, mais décider ensuite de me procurer un énorme gâteau au chocolat. Il peut s'agir d'un cas de faiblesse de volonté, même si, au moment venu, j'échoue à obtenir un gâteau au chocolat ou que je n'accomplisse pas l'action nécessaire pour l'obtenir. À partir du moment où je décide d'adopter une ligne de conduite allant à l'encontre de mon meilleur jugement préférentiel, je suis réputé vivre un épisode de stricte akrasie.

Il est trivial de dire qu'un agent rationnel doit élaborer des scénarios réellement praticables par lui. Il ne sert à rien de délibérer sur des alternatives qu'on ne peut pas de toute façon choisir. Toutefois, on risque de négliger cette lapalissade si l'on se contente de (CS) sans clauses limitatives ou précisions supplémentaires. Je peux prendre une décision contre-préférentielle ou à l'encontre de mon meilleur jugement parce que je suis finalement incapable de respecter un plan que je me suis pourtant donné. Cela n'implique évidemment pas que je vis un épisode de faiblesse de volonté. Mais on peut oublier cette limitation si

notre attention est portée exclusivement sur les facteurs physiques qui sont susceptibles de contraindre la réalisation de nos plans et décisions pour l'avenir.

En effet, les contraintes dont on doit tenir compte dans la planification rationnelle de nos activités ne doivent pas se réduire à des contraintes externes comme l'interférence d'un tiers ou la non-réalisation des conditions minimales de réalisation, ni à des contraintes internes comme nos capacités motrices. Il est vrai que si j'estime qu'il est préférable de me rendre à une soirée pour 21 h et que le meilleur moyen est que je coure à 50 km/h pour m'y rendre, je ne vivrai pas un épisode de faiblesse de volonté si je n'arrive pas à maintenir cette vitesse ou même si je ne l'atteins pas. J'ai tout simplement mal évalué au départ mes capacités. Souvent, notre connaissance intime et tacite de nos propres capacités motrices filtre en amont les alternatives que l'on va considérer dans notre délibération. Ce filtre fonctionne souvent correctement, mais il peut arriver qu'il laisse passer trop ou trop peu. Beaucoup de blessures, par exemple, s'expliquent par une surestimation de nos capacités motrices[38]. Cela est également vrai des contraintes internes qui sont de nature cognitive. Des capacités cognitives comme la mémoire de travail et la mémoire épisodique, le bagage informationnel dont nous disposons, comme le savoir factuel et procédural, les habiletés logiques et les capacités d'attention sont foncièrement limitées. Aussi, l'agent qui les ignore s'expose à des risques d'échecs importants.

Ce dernier fait a tendance à passer inaperçu lors de diagnostic de cas de faiblesse de volonté simplement parce qu'on suppose d'emblée que les agents ont déjà filtré en amont les alternatives qui sont impraticables pour eux. En fait, ce filtre ne fonctionne pas toujours avec une grande précision. Les agents ont tendance à surestimer leurs capacités cognitives – en particulier la fiabilité de leurs mémoires et le maintien de leur attention. Aussi, ils sont plus prompts à s'engager dans des lignes de conduite qu'ils ne pourront pas tenir longtemps. Autrement dit, les agents ont tendance à élaborer des plans irréalistes du point

[38] Nos limites motrices sont appréciées au moyen de situations-tests comme les jeux et les sports, et l'expérience que nous en avons depuis la tendre enfance. Mais ces limites sont également appréciées par le biais de dispositifs proprioceptifs et intéroceptifs qui nous permettent de connaitre, soit en temps réel soit de manière différée, l'état de notre charpente musculo-squelettique, et l'état de nos viscères. Les personnes atteintes d'anosognosie ont des handicaps moteurs importants (paralysie d'un membre ou de tout un côté du corps), mais nient catégoriquement les avoir. Des lésions du cerveau dues souvent à une attaque cérébrale altèrent significativement de manière irréversible le fonctionnement de ces dispositifs. Il en résulte un trouble des plus étranges. Même si les malades observent que leur membre ne bouge pas, ils continuent d'affirmer sincèrement qu'ils ont conservé les mêmes capacités motrices qu'auparavant. Il ne s'agit pas d'un problème psychologique de croyance motivée ou d'un phénomène de réduction de la dissonance cognitive, mais d'un problème neurologique touchant des fonctions fondamentales d'autorégulation du corps. (Damasio, 1995, 2010)

de vue cognitif, et cela peut constituer un problème dans le diagnostic de cas de faiblesse de volonté.

Une autre raison de réviser en ce sens (CS) est que cela permet de considérer des cas où l'agent agit en accord avec son meilleur jugement, mais semble néanmoins vivre un épisode de stricte akrasie. Je peux estimer qu'il est préférable, toutes choses considérées, de ne pas sauter en parachute ce weekend. Or, pris d'un enthousiasme débridé, je me rends au club de parachutisme et m'inscris dans la liste des passagers-sauteurs du premier vol de la journée de samedi. Au moment où je dois embarquer dans l'avion, je suis si transi de peur que je ne me résous pas à monter à bord. Aussi, je n'agis pas à l'encontre de mon meilleur jugement. Par contre, j'ai clairement pris une décision à l'encontre de mon meilleur jugement quand j'ai choisi de me rendre au club dans l'espoir de sauter en parachute[39].

Cela dit, contrairement aux économistes et psychologues, les philosophes sont plus enclins à parler d'*intention* que de *décision* pour expliquer les comportements des agents. Auquel cas, il est peut-être plus approprié d'utiliser cette dernière notion pour définir la faiblesse de volonté. Les agents sont réputés akratiques lorsqu'ils *forment l'intention* d'agir à l'encontre de leur meilleur jugement. Dans ce contexte, les deux expressions ont le même sens. Par contre, il faut distinguer l'*intention d'agir* et l'*intention dans l'action*. L'intention de faire *x* équivaut logiquement à décider de faire *x*. Mais faire *x* intentionnellement n'équivaut pas à avoir décidé de faire *x*. Faire quelque chose intentionnellement peut vouloir dire qu'on a pris la décision de le faire ou peut vouloir dire que lorsqu'on l'a fait, on connaissait certaines conséquences pertinentes qui n'étaient pas voulues ou visées. Demander à quelqu'un s'il a, par exemple, intentionnellement crevé l'œil du chasseur qui se tenait près du gibier lorsqu'il tira dans la direction de ce dernier, c'est lui demander s'il estimait, au moment où il a pris la décision de tirer, que cela comportait un risque important pour l'œil du chasseur, ou s'il estimait qu'il n'y avait pas de tels risques. S'il estimait qu'il y avait un tel risque, alors on dira qu'il a intentionnellement crevé l'œil du chasseur, même si cela ne

[39] Cela, je crois, correspond à la manière habituelle de diagnostiquer un problème de faiblesse de volonté. Chacun estime qu'un ex-toxicomane ou un ex-alcoolique vit une rechute au moment où il décide de consommer des substances addictives et non pas simplement au moment où il en consomme effectivement. Aussi, la période à risque, où tout peut basculer pour eux, ne se situe pas entre la décision de réintégrer leurs anciennes habitudes et le point de consommation, mais entre les premières sensations de manque et la décision de combler ce manque. Une fois la décision prise, il est souvent trop tard. C'est pourquoi des groupes de soutien comme les Alcooliques Anonymes concentrent leurs stratégies d'interventions critiques sur la période d'indécision qui précède une rechute possible.

constituait pas pour lui un résultat visé, ni même simplement souhaité – peut-être souhaitait-il seulement maximiser ses chances d'abattre le gibier et jugea le risque qu'il faisait courir à l'autre chasseur comme le prix à payer. Dans cette situation, on ne dira pas que l'agent avait l'intention de crever l'œil du chasseur ou qu'il a pris la décision de lui crever l'œil, et cela, même si l'agent estimait que le fait que le chasseur ait l'œil crevé avait quelque chose de désirable. Bien que le terme « décision de faire *x* » ait la même extension que le terme « avoir l'intention de faire *x* », je continuerai néanmoins à utiliser la notion de décision, en dépit du fait que la notion d'intention est privilégiée dans la littérature philosophique, parce qu'elle ne charrie pas autant d'ambiguïtés.

Faire référence aux actions est aussi un facteur de confusion. Agir à l'encontre de son meilleur jugement peut être attribuable à un problème moteur et non motivationnel. Avec la notion de « décision » (ou choix effectif, acte de choix, formation d'une intention), on évite ce genre de confusion puisqu'une décision est un événement mental qui se situe entre le résultat de la délibération (formation du jugement) et son exécution (action)[40]. Aussi, un agent peut souffrir d'un problème moteur dans l'exécution de ses décisions tout en faisant des choix rationnels et en restant résolu.

Si ce qui précède est correct, cela nous amène donc à réviser la définition standard de manière suivante

Conception Standard Révisée (1) :

(CSR1) : *En décidant de faire* x*, un agent fait un choix incontinent, si et seulement si : a) l'agent croit qu'il peut aussi choisir de faire* y*; et b) l'agent juge que, tout bien considéré, il serait mieux de choisir* y *que de choisir* x.

Ici la clause (a) de (CS) a disparu dans la version révisée parce que superflue. Comme on ne peut pas décider de faire quelque chose de manière non intentionnelle, on peut se passer de cette clause[41]. Je pense que cette définition est intuitivement plus adéquate tout en

[40] Voir Holton (2006) pour une défense de l'hypothèse que les décisions sont des événements mentaux distincts des croyances et des désirs, qui jouent un rôle important dans l'explication des actions.

[41] Par contre, pour que notre définition fonctionne, on doit semble-t-il stipuler une contrainte de précision concernant la manière de décrire les décisions akratiques. Si, par exemple, je forme le jugement qu'il est préférable pour moi de cesser ma consommation de crack, mais décide ensuite de me rendre chez mon ancien « *dealer* », ma décision de me rendre chez lui n'est pas à proprement parler un cas de décision akratique. Mais si cette décision est motivée par le fait que je pourrai m'y injecter une dose, alors il sera plus approprié de décrire celle-ci comme une décision de consommer du crack, auquel

respectant l'esprit de la conception standard (CS). Le lecteur peut penser qu'une révision de la sorte de la définition standard est somme toute mineure et ne requiert pas d'être mentionnée. Il est vrai que cela ne semble pas changer significativement la manière d'interpréter les positions et théories philosophiques sur l'akrasie. Or, je pense que la (CSR1) est pertinente. Comme je l'ai mentionné dans l'introduction, on peut distinguer clairement les conceptions de la faiblesse de volonté identifiant le problème dans le processus qui *précède* la décision akratique des conceptions identifiant le problème dans le processus qui *suit* cette décision.

On peut également juger judicieux de réviser la (CS) en y ajoutant des clauses temporelles qui spécifient les instants ou les moments où l'akratès prend une décision, croit et estime qu'une ligne de conduite est préférable à une autre. En spécifiant correctement ces instants ou ces moments, on évite de confondre la stricte akrasie avec une simple inversion des préférences. Une des erreurs qu'il est facile de commettre est de considérer que les agents ont des préférences suffisamment stables dans le temps pour qu'on soit tenté de traiter leurs jugements préférentiels comme des préférences *a*temporelles. Cela a pour effet de multiplier les cas de stricte akrasie. Lorsqu'on constate, par exemple, qu'un agent estime qu'il est préférable de ne pas fumer, on conclut *ipso facto* qu'il accomplit un acte de stricte akrasie s'il s'allume une cigarette[42]. Or, il y a toutes sortes de facteurs externes, mais aussi internes qui permettent d'expliquer pourquoi des agents, au premier abord rationnels, modifient sans crier gare leurs préférences. Leurs préférences peuvent être affectées en profondeur par des stimuli tels que des déclencheurs perceptuels ou cognitifs ainsi que par des motivations viscérales comme décrites dans le chapitre 1. En ajoutant des clauses temporelles à la définition pour éviter de confondre la stricte akrasie avec un problème d'instabilité dynamique des préférences, on obtient quelque chose comme :

---

cas je déciderais clairement à l'encontre de mon meilleur jugement. Du reste, les problèmes de précision dans l'identification des décisions font écho aux problèmes du choix de la bonne description sous laquelle une action est intentionnelle pour un agent. En fait, le problème est identique si l'on considère, comme je le crois, qu'une décision de faire *x* correspond à l'intention de faire *x*, et que prendre la décision de faire *x* revient à former l'intention de faire *x*.

[42] C'est ce qui expliquerait, selon Joseph Heath, l'état de confusion qui règne dans le milieu philosophique à propos du fait qu'il existerait quelque chose comme la stricte akrasie (2008 : 228-234).

Conception Standard Révisée (2) :

(CSR2) : *En décidant à* t *de faire* x*, un agent fait un choix incontinent si et seulement si : a) l'agent croit à* t *qu'il peut aussi choisir de faire* y*; et b) l'agent juge à* t *que, tout bien considéré, il serait mieux de choisir* y *que de choisir* x.

En indexant les jugements préférentiels et les décisions au temps, on prépare mieux le temps pour un débat sur la possibilité de la stricte akrasie et la nature de la faiblesse de volonté en général. C'est pourquoi j'y ferai toujours référence, si ce n'est qu'implicitement, lorsque je discuterai des arguments sceptiques et ceux en faveur de la stricte akrasie. Mais examinons maintenant en quoi consistent les jugements préférentiels des agents, puisqu'ils jouent un rôle si important, selon la conception standard ((CS) ou (CSR1-2)) dans le diagnostic de cas d'akrasie.

La conception Standard formulée initialement par Davidson met l'emphase sur une incohérence entre un jugement de l'agent et son action (ou décision). Il s'agit à mon avis d'une conception un peu trop restrictive de la stricte akrasie. Si l'on suppose, au préalable, que les agents *rationnels* agissent ou décident en fonction de *jugements*, la conception est alors pleinement opérationnelle – c'est-à-dire qu'on peut s'en servir pour identifier tous les cas de stricte akrasie. Mais si l'on estime plutôt que les agents rationnels agissent ou décident en fonction de *préférences* (désirs, goûts, appétits, espoirs, convoitise, penchant, etc.), la conception perd quelque peu de sa portée. Intuitivement parlant, on peut considérer des cas de décisions contre-préférentielles, *et sur cette seule base*, comme des cas putatifs de stricte akrasie. Je pourrais dire que je préfère, tout bien considéré, ne pas fumer, mais décider tout de même d'acheter un paquet de cigarettes.

3.1.3 Deux traditions en théorie de la motivation et l'inclusion de la clause « estime que… »

Ce qui précède peut sembler être un point relativement mineur. Mais comme il y a deux traditions théoriques concernant la nature des constituants ultimes de la motivation, on doit en tenir compte lorsqu'on élabore une conception de la stricte akrasie.

Il y a la tradition théorique qui fait des jugements les constituants fondamentaux de la motivation. Comme on l'a vu cette tradition remonte au moins à Platon et à Aristote. Dans cette tradition, les considérations désidératives et affectives relèvent de simples « inputs »

dans le processus délibératif. C'est au final le jugement de l'agent qui assigne des poids à des considérations en fonction idéalement de leur importance réelle et motive la décision. Il s'agit clairement d'une forme de cognitivisme motivationnel. Dans ce cadre, la stricte akrasie apparait paradoxale. Elle devient possible seulement si le jugement peut à l'occasion échouer à se connecter au circuit motivationnel de l'agent parce qu'une cause plus forte – de nature irrationnelle – l'empêche de le faire ou prend le relais.

La tradition fondée sur les préférences, dont les théories hédonistes en sont les principaux représentants, fait des considérations désidératives (et dans certains cas affectives) les constituants fondamentaux de la motivation. Par le fait même, elle fait jouer un rôle plutôt périphérique aux jugements. La première version un tant soit peu élaborée d'une conception hédoniste de la motivation remonte à Bentham. Pour Bentham,

> « […] *nature has placed mankind under the governance of two sovereign masters,* pain *and* pleasure. *It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne. They govern us in all we do, in all we say, in all we think: every effort we can make to throw off our subjection, will serve but to demonstrate and confirm it.* » (1789 : § 1)

Dans son ouvrage classique, *Introduction to the Principles of Morals and Legislation***,** Bentham a décrit les diverses sources de plaisir et de douleur qui enclenchent des réponses d'approche ou d'évitement. Il avançait néanmoins l'idée que l'ensemble des motifs de choix est synthétisable dans un seul et unique motif qu'il appela l'« *utilité* ». La notion d'utilité fut ensuite largement adoptée par les théoriciens en philosophie morale, en théorie économique du choix, en psychologie comportementale et par les sciences cognitives. La nature de l'utilité dans les théories à saveur hédonique varie toutefois énormément. Comme Kahneman l'a fait remarquer, la notion d'utilité telle que Bentham la concevait concerne plus les états et les processus affectifs engendrés par des conséquences de ses choix que de la désirabilité *stricto sensu* de ces mêmes conséquences (2006 : 489). Cette dernière interprétation de la notion est plus conforme aux théories et aux dispositifs conceptuels élaborés dans la science économique moderne – du moins depuis les travaux fondateurs de l'économiste Irving Fisher (1916 : 49-51).

Beaucoup de théories et d'approches contemporaines en neurologie, en sciences cognitives et en science comportementales acceptent l'idée que des états hédoniques sont l'horizon ultime de nos choix, mais que la qualité de délibération ou de la réflexion peut tout de

même faire la différence (Dawes & Hastie, 2010). Je ne discuterai pas ici davantage des mérites de l'approche hédonique ou à saveur hédonique de la motivation. Comme je l'ai mentionné dans le 1er chapitre, assigner des poids à des préférences est un processus lui-même déterminé par des poids[43]. Ce que je veux mettre en relief ici est le fait qu'une définition standard de la stricte akrasie doit être suffisamment neutre pour être interprétable dans le paradigme des deux traditions de la motivation – même si la tradition hédonique est conceptuellement et empiriquement supérieure.

En fait, si les agents n'agissent pas ou ne décident pas par jugements (considérations cognitives), mais par préférences (considérations désidératives), alors la stricte akrasie, *telle que décrite par la conception Standard*, ne pose aucun problème théorique; comme les décisions sont motivées par autres choses que des jugements, il n'est pas paradoxal que les agents décident à l'encontre. En revanche, si l'on substitue dans la définition Standard la clause « … *préfère que* ... » à « … *juge que*... », et qu'on prétend que les agents choisissent par préférence, alors la stricte akrasie garde son caractère paradoxal.

Dans les deux cas, cependant, on a affaire à des décisions non-maximisantes qui engendrent systématiquement des résultats sous-optimaux. Dans le cadre de tradition hédonique, l'agent akratique échoue à maximiser son bien-être conçut comme un état hédonique de satisfaction, tandis que dans le cadre de la tradition cognitive, il échoue à maximiser son bien-être conçu cependant comme des états non affectifs, comme l'eudémonia aristotélicienne, par exemple, ou le salut chrétien. On peut, je pense, reformuler l'énoncé de la conception Standard en remplaçant la clause « ... *juge que*... » par la clause « …*estime que*... ». Cette dernière est beaucoup plus neutre à l'égard de la nature des éléments qui forment le soubassement de la motivation :

Conception Standard élargie :

(CSE) : *En décidant à* t *de faire* x*, un agent fait un choix incontinent, si et seulement si : a) l'agent croit à* t *qu'il peut aussi choisir de faire* y*; et b) l'agent estime à* t *que, tout bien considéré, il serait mieux de choisir* y *que de choisir* x.

[43] Il serait plus précis et plus correct de dire ici que l'assignation de poids dans le processus de décision est *sous-déterminée* (mais déterminée tout de même) par des poids. Comme je le montrerai dans le chapitre 7, les agents rationnels utilisent consciemment ou non toute une série de règles et de procédures décisionnelles variées qui les amènent à faire des choix différents à partir d'un même stock initial de préférences.

Bien que le verbe « estimer » ait une connotation plus cognitive que désidérative, il demeure assez neutre pour décrire l'attitude évaluative que les agents ont à l'égard des alternatives disponibles, sans spécifier la nature des éléments de base de cette attitude. On peut, dans un cadre plus technique et conforme aux théories économiques modernes, considérer que lorsque l'agent estime qu'il serait mieux de choisir *y* que de choisir x, il consulte une sorte de menu interne qui fait état des utilités associées à la réalisation des alternatives en lice. Mais on peut interpréter la clause « … *estime que*… », dans une optique plus traditionnelle, comme une forme de jugement. Aussi, dans ce qui va suivre, je me référerai à la version (CSE) plus neutre lorsque je discuterai de la conception Standard.

3.2 Conclusion

Le rôle d'une conception Standard de la stricte akrasie est celui d'un point de référence pour toute discussion de sa nature et de sa possibilité. Une telle conception doit être suffisamment précise et neutre pour éviter les confusions et nous orienter correctement dans l'enquête philosophique et scientifique, mais aussi dans les débats et les disputes qu'elle suscite. Puisqu'elle fut au cœur de beaucoup de discussions et qu'elle est suffisamment claire, la conception de Davidson est la meilleure candidate pour jouer ce rôle. Elle doit, par contre, subir quelques modifications pour être suffisamment précise et générale. Il est plus approprié qu'elle porte sur des décisions que sur des actions. Mais surtout, elle ne doit pas préjuger de la manière dont les agents procèdent pour décider, s'ils le font à partir de jugements ou à partir de désirs ou de préférences.

Avoir en main une conception Standard n'a pas que des avantages. Les disputes philosophiques auxquelles une conception de référence peut donner lieu peuvent trop orienter l'attention sur la conception elle-même que ce sur quoi elle est censée porter. Cela peut donner l'impression que le débat porte sur un problème créé de toutes pièces par les philosophes et que ceux-ci oublient le problème réel que pose la faiblesse de volonté pour la rationalité pratique (ex. : pourquoi je continue à fumer alors que cela m'empoisonne, ou pourquoi je fais toujours mon rapport d'impôt après la date limite? Ou pourquoi je ne peux m'arrêter de dépenser quand je vais dans les boutiques? etc.). Aussi, on doit toujours garder à l'esprit que les disputes sur la possibilité de la stricte akrasie sont en fait des disputes sur la motivation des agents.

Avec la conception Standard (CSE) en guise de référence, j'aborderai la théorie des préférences révélées. Il s'agit d'une théorie plutôt austère de la motivation qui a beaucoup d'adhérents déclarés – surtout parmi les chercheurs en microéconomie –, mais également beaucoup de non déclarés. Si la conception Standard (CSE) est correcte, la théorie des préférences révélées implique *nécessairement* l'inexistence de la stricte akrasie. Il s'agit donc d'une position sceptique forte, et à ma connaissance la plus radicale. Pour toutes ces raisons, le prochain chapitre lui sera consacré.

4

La théorie des préférences révélées : une position sceptique radicale

Deux économistes contemplent une Ferrari lorsque l'un d'eux dit : « J'aimerais en avoir une ». L'autre rétorque aussitôt : « Ce n'est manifestement pas le cas! ».

–Blague populaire parmi les économistes

4.1 La proposition de base de la théorie

Pour beaucoup de théoriciens, la conception Standard de la faiblesse de volonté décrirait un phénomène si paradoxal que d'aucuns verront en lui une sorte de chimère conceptuelle. Les raisons qui motivent ce scepticisme sont par contre très variées. Dans le domaine philosophique, beaucoup estiment qu'il s'agit d'une impossibilité logique, conceptuelle ou même dont l'existence irait à l'encontre des règles sémantiques qui encadre l'usage des termes « décision », « intentionnel », « préférences » ou « jugement » (Heath, 2008). Mais il existe également des considérations méthodologiques et épistémiques pour nier l'existence du phénomène. Je réserve le traitement des premiers arguments au chapitre 6 et me concentrerai ici sur l'examen d'un argument méthodologico-épistémique.

Comment peut-on nier l'existence d'un phénomène comme celui décrit par la conception Standard sur une base méthodologique et épistémique? La réponse vient non pas d'une théorie philosophique, mais d'une théorie économique du choix rationnel. Il s'agit de la *théorie des préférences révélées*. Formulée initialement à la fin des années 1930 par l'économiste Paul Samuelson (1938), et raffinée par la suite par Hendrik Houthakker (1950), cette théorie traduit un souci méthodologique clair : il est inutile de faire des hypothèses sophistiquées sur les préférences des agents, quand on peut les déduire directement et exclusivement des choix ou décisions que ces derniers prennent effectivement[44]. Simple, élégante et hautement systématisée, la théorie des préférences révélées est constituée par une série d'axiomes censés initialement expliquer le choix des

[44] Voir également la description et la critique qu'en fait Amatya Sen (2005 : 103-109).

consommateurs à l'égard de différents paniers de biens. Je n'ai pas l'intention ici de présenter ces axiomes. Ils sont abstraits et ne sont pas particulièrement intéressants pour l'enquête philosophique. Bien qu'il y ait plusieurs versions de cette théorie, j'aborderai seulement une proposition fondamentale qui forme le cœur de cette théorie. On peut formuler succinctement cette proposition sous forme du principe méthodologique suivant :

Principe des préférences révélées :

(PPR) : *Un agent estime à* t *qu'il est préférable de faire* x *plutôt que* y *si et seulement si : a) l'agent croit à* t *qu'il peut aussi choisir de faire* y *; b) l'agent décide à* t *de faire* x.

(PPR) est une proposition qui présentait initialement un attrait important. Elle exprimait un souci méthodologique qui s'inscrivait alors dans la vague béhavioriste de l'époque, laquelle manifestait un dénigrement des concepts mentaux jugés incompatibles avec l'enquête scientifique rigoureuse. D'ailleurs, la science économique de l'époque aspirait à produire des explications autonomes qui ne sont tributaires d'aucune autre science – si ce n'est la science des mathématiques. On laissait aux sociologues l'étude des comportements irrationnels et l'on cherchait à dépsychologiser les notions de microéconomie. La théorie des préférences révélées dénote notamment ce souci de dé-psychologisation.

La théorie présente également une valeur intuitive importante. Nous faisons régulièrement des hypothèses sur les préférences, les désirs, les goûts, les souhaits ou les espérances d'autrui sur la base du simple examen de ses choix. Il nous arrive même à l'occasion de penser que nos hypothèses sont plus fiables et robustes que les rapports introspectifs qu'autrui nous présente ou se présente à lui-même.

Il n'est pas difficile de voir, si l'on accepte cette proposition, qu'on est obligé de nier la possibilité de la faiblesse de volonté. Seulement, il n'est pas vrai que la conjonction de (PPR) et de (CSE) entraîne à elle seule l'inexistence de cas de stricte akrasie. On doit également inclure l'hypothèse que les agents sont des maximisateurs (d'utilités espérées, de plaisir ou de bien-être). En effet, on peut bien dire que les décisions des agents reflètent leurs préférences, mais on doit en plus spécifier laquelle. Avec l'hypothèse de la maximisation, on est forcé de dire que ce sont les préférences *dominantes* des agents qu'ils satisfont en choisissant. Sans cette hypothèse, on pourrait déduire, de l'observation que

Paul a choisi une pomme plutôt qu'une poire, qu'il a une préférence plus grande pour la poire. Or, l'hypothèse de la maximisation exclut ce genre d'inférence. Si Paul a choisi la pomme, c'est nécessairement parce qu'il a une préférence plus grande pour ce fruit que pour l'autre. Aussi, la conjonction de (PPR), de (CSE) et de l'hypothèse de la maximisation entraîne la conclusion qu'il n'y a pas de décision qui satisfait (CSE). Comme les jugements préférentiels des agents sont exclusivement reflétés dans les choix qu'ils font, alors ils ne peuvent pas prétendre préférer une autre ligne de conduite que celle qu'ils adoptent, parce qu'ils sont des maximisateurs d'utilité. Je peux bien estimer qu'il serait préférable maintenant que je ne fume pas, mais comme je fume, il s'ensuit que je préfère fumer.

L'hypothèse de la maximisation n'est pas qu'une simple hypothèse auxiliaire nécessaire à la théorie des préférences révélées pour la dérivation d'une conclusion sceptique. Cette hypothèse, qui remonte à Platon, est en fait au cœur de toutes les positions sceptiques qu'on trouve sur le marché philosophique et sur celui des sciences comportementales. J'estime qu'il s'agit d'une hypothèse erronée, mais je compte en faire seulement la critique au chapitre 7. Pour l'instant, examinons la nature des préférences que la théorie voit à la surface des choix effectifs des agents et la portée de la théorie pour expliquer les décisions des agents – tant dans des contextes moraux que prudentiels.

4.2 La nature des préférences et la grande portée de la théorie

La notion de préférence dans la théorie des préférences révélées ne correspond pas parfaitement à la notion qu'on utilise dans le langage ordinaire. Il s'agit d'un concept technique qui est une extension de la notion économique d'utilité espérée, laquelle est en quelque sorte la version numérique (sur une échelle cardinale) de la première (sur échelle ordinale). La différence entre la notion ordinaire et la notion technique est analogue à celle entre la notion ordinaire de poids et celle de masse. Ces notions ne correspondent pas tout à fait, mais entretiennent une relation de proche parenté. Je vais me tenir ici à une version non technique de la notion parce que cela n'apporterait rien de significatif à la discussion que d'utiliser une notion trop technique.

4.2.1 Le caractère évaluatif des préférences

Les préférences reflètent la valeur que les choses ont pour nous. La notion de préférence est un terme générique qui embrasse toutes sortes de considérations désidératives comme les désirs, les souhaits, les penchants, les envies, les espoirs, les convoitises, les appétits, les ambitions, les caprices, mais aussi des aversions corrélatives comme des peurs, des répugnances, des appréhensions, des répulsions, etc. Les préférences peuvent être très idiosyncrasiques comme le sont les désirs et les goûts pour telle ou telle organisation de notre jardin ou pour les sortes de bières ou plus impersonnelles et transculturelles comme les préférences pour la santé, la sécurité et les opportunités. Au niveau de l'économie de base des organismes vivants, les préférences correspondent aux valences de leurs états internes. Comme les organismes tentent de maintenir un état homéostatique stable situé dans une fourchette préalablement fixée, les ressources externes seront évaluées en fonction du maintien de l'état homéostatique (Damasio, 1995). Pour des agents rationnels sophistiqués, les préférences correspondent à des *assignations subjectives de valeur* à des options qui peuvent avoir un lien très indirect ou inexistant avec l'état homéostatique de l'agent. Ces valeurs peuvent être corrélatives de préférences très abstraites (ex. : l'espérance de la réalisation d'une utopie) ou très concrètes (ex. : l'appréciation anticipée d'une tarte au sucre). On peut en faire l'expérience de manière viscérale, comme lorsqu'on a un penchant, une envie ou une haine, mais nos préférences peuvent aussi passer sous le radar mental et s'avérer plutôt discrètes, comme nos préférences pour un environnement sain ou pour le respect des règles de bonne gouvernance dans nos institutions.

Les options comme telles, objets des préférences, doivent être des *alternatives* et non des choses qu'on peut obtenir ou des actes qu'on peut accomplir de manière non mutuellement exclusive. On n'a pas à choisir si l'on ira au cinéma samedi ou au restaurant s'il n'existe pas de considérations externes qui nous forcent en quelque sorte à renoncer à l'une de ces options. Je peux me donner un budget de sortie qui ne me permet pas d'aller au restaurant et au cinéma la même semaine.

L'assignation des valeurs au travers des décisions des agents produit un *ordonnancement* (*ranking*) des alternatives en fonction de leur valeur. Si je considère les options d'achat [*a, b, c*] dans une boutique, on obtiendra un tableau ou un menu de mon ordonnancement [$a > b > c$], si et seulement si je choisis *a* alors qu'on me propose en plus les alternatives *b* et *c*,

et *b* alors qu'on me propose *c*. Mais dans chacun des cas, mon choix devrait révéler une préférence dominante pour une alternative. Si je me montre indécis cependant, cela est censé montrer qu'il n'y pas une préférence dominante pour une des options dans la série qui s'offre à moi, mais une indifférence entre au moins deux options qui ont pour moi la même valeur.

4.2.2 Des choix aux préférences : une relation analytique

Dans la définition Standard (CSE), on mentionne le fait que l'agent akratique estime, *toutes choses considérées*, qu'une option est meilleure qu'une autre. Or, il n'est pas nécessaire de mentionner cela si l'on adhère à la théorie des préférences révélées. Dans le cadre de cette théorie, on n'a pas besoin de savoir comment les agents ont établi par comparaison leur ordonnancement des alternatives, puisqu'il se déduit complètement des choix effectifs.

Par contre, cela n'implique pas qu'on ne puisse pas investiguer et donner des détails sur les processus qui ont mené à leur formation. Car, si c'était le cas, la théorie des préférences révélées serait un dispositif théorique stérile qui coupe l'herbe sous le pied de quiconque cherche à découvrir les motifs fins et subtils qui expliquent quelques fois les décisions des agents. Seulement, on doit toujours *partir* des choix et identifier ensuite les préférences aux couches supérieures, lesquelles peuvent être expliquées par d'autres motifs ou préférences de niveau inférieur, etc. L'investigation des motifs doit suivre un schéma analytique et non synthétique.

La Théorie du choix rationnel en économie fait état de préférences partielles (ou critères partiels) à la base des préférences complètes (*critères complets*) situées à la jonction des décisions (*at the point of decision*). De même, on trouve dans la littérature philosophique, depuis la parution du célèbre article de Davidson sur la possibilité de la faiblesse de volonté, des références aux jugements partiels[45]. Cela reste compatible avec la théorie des

[45] Les préférences partielles peuvent être exprimées par d'autres types de jugement. On peut les exprimer notamment par des jugements *pro tanto* qui prennent la forme canonique : « *Dans la mesure où P, x est préférable à y* ». Expliquer pourquoi on juge, *pro tanto*, qu'une option est préférable à une autre implique qu'on identifie une propriété relationnelle qui *ferait* la différence si certaines conditions étaient réalisées. Mais on peut également exprimer des préférences partielles par des jugements *ceteris paribus* qui prennent la forme canonique : « *Toutes chose étant égales par ailleurs, x est préférable à y*». Quand on explique pourquoi, *ceteris paribus*, *x* est préférable à *y*, on identifie habituellement une propriété qui *ferait* la différence si toutes les autres étaient identiques.

Mais la littérature philosophique voit typiquement les préférences partielles comme des jugements *prima facie*, depuis Davidson (Davidson : 1970). Il est naturel de le faire parce que les jugements partiels peuvent être considérés comme des

préférences révélées. Différents jugements préférentiels partiels peuvent orienter les agents dans des directions différentes et incompatibles sans qu'ils soient eux-mêmes logiquement incompatibles. Les préférences situées à la jonction des décisions doivent souvent refléter une solution de compromis (*trade-off*) entre l'ensemble des diverses considérations[46].

Thierry peut opter pour Clarisse même si Clara a une chevelure plus abondante et qu'il préfère *de manière partielle* (*pro tanto*, *prima facie*) les femmes qui ont des chevelures abondantes, et ce choix n'a rien d'énigmatique ou de paradoxal. Lorsqu'on souhaite mettre en relief le caractère paradoxal de la faiblesse de volonté, on ne peut pas le faire en mentionnant seulement des jugements préférentiels partiels. Étant données les préférences de Thierry pour les chevelures abondantes, on ne dira pas qu'il vit un épisode de faiblesse de volonté s'il choisit de s'acoquiner avec une femme chauve. Par contre, on dira qu'il vit un tel épisode – conformément à (CSE) – s'il agit à l'encontre d'une préférence située à la jonction de sa décision. Or, cela est clairement exclu par la théorie des préférences révélées.

4.2.3 Les contenus égoïstes/altruistes des préférences

Dans la volonté des philosophes, économistes et politicologues d'après-guerre d'établir des conceptions plus « réalistes » de la motivation humaine et une théorie plus systématique de la rationalité pratique – reposant notamment sur des principes axiomatisés – et surtout dénués de toutes références aux attitudes « *sentimentaliste* » des personnes irrationnelles,

---

résultats d'étapes dans un processus de délibération visant à identifier le choix le plus éclairé ou les plus avantageux. Je ne nie pas d'ailleurs qu'il y ait un usage de l'expression *prima facie* qui soit étranger au résultat d'étape dans la délibération rationnelle, et qui ait plus à voir avec l'identification d'apparences trompeuses que de propriétés réelles ou escomptées que les options sont censées exemplifier. *Prima facie*, ce modèle automobile est plus fiable que l'autre. Mais après une recherche approfondie, il s'avère que ce modèle est en fait beaucoup moins fiable que l'autre. On peut s'apercevoir qu'on s'est laissé berner par le caractère rutilant des pièces apparentes, par exemple. Ici le jugement *prima facie* ne doit pas être considéré lors de la délibération rationnelle – ce qui n'est du reste pas le cas de jugements *prima facie* qui identifient réellement des propriétés que l'agent cherche à maximiser ou minimiser.

Cette confusion des usages se reflète notamment dans la critique qu'a fait Kagan de l'usage de l'expression « prima facie » dans la théorie morale de David Ross (1930). Ross a défendu l'idée que la morale doit reposer sur un corpus de devoir *prima facie*, c'est-à-dire des devoirs auxquels nous devons, intuitivement, nous conformer à moins qu'il existe un autre devoir dont nous avons l'intuition qu'il prévaut dans la situation sur le premier. Kagan suggère de remplacer dans ce contexte « prima facie » par « pro tanto », alléguant qu'une raison *pro tanto* doit toujours avoir du poids dans la délibération, même si c'est moins qu'on le pensait au début, alors qu'une raison *prima facie* a typiquement du poids au début de la délibération, mais aucun dans vers la fin (1989 : 17).

Je n'utiliserai pas l'expression « prima facie » dans le sens de Kagan, mais je me conformerai à l'usage philosophique qu'elle a en théorie de l'action et du choix rationnel, et non à l'usage qu'elle a dans les domaines métaphysique ou d'épistémologie fondamentale.

[46] Dans le cadre de la théorie, on peut théoriquement isoler chacune de ces considérations en demandant à l'agent de faire des choix dans d'autres contextes en modifiant qu'une seule variable à la fois. C'est ce qu'on identifie hypothétiquement lorsqu'on utilise la clause « *ceteris paribus* ».

on a cru bon d'exclure les préférences altruistes dans le calcul rationnel. Une forme de « réalisme » extrême en vogue à l'époque affirmait que les agents sont exclusivement motivés par des considérations relevant, ne serait-ce qu'indirectement, de leurs intérêts personnels et que, par conséquent, les considérations morales ne sont pertinentes d'un point de vue explicatif que dans la mesure où elles leur procurent du plaisir ou améliorent leurs conditions d'existence. Bien que cette hypothèse aille à l'encontre de certaines de nos intuitions les plus fortes, il reste indéniable que certains auteurs ont su débusquer des comportements intéressés là où il semblait y avoir pur altruisme. On trouve d'ailleurs dans la littérature romanesque des descriptions fines de l'attitude de personnages qui éprouvent un plaisir vif dans la charité et l'abnégation, au point où il est difficile de dire que leur comportement est réellement désintéressé.

Or, bien que la théorie des préférences révélées ait été initialement élaborée en économie pour expliquer le choix des consommateurs, il n'est pas conceptuellement nécessaire que les préférences doivent avoir un contenu de nature strictement égoïste. Nous n'avons pas besoin d'adhérer au modèle de l'*homo oeconomicus* pour accepter la théorie. En fait, rien ne s'oppose conceptuellement à l'idée que des préférences puissent porter sur des options qui produisent un impact hédonique positif supérieur – voire exclusif – sur autrui. Cela laisse une place à l'explication d'une classe importante de comportements moraux à l'intérieur du cadre conceptuel de la théorie des préférences révélées, et permet de statuer sur les cas de faiblesse de volonté et d'accidie morale.

Il existe toute une littérature sur le problème de l'altruisme véritable et je n'entends pas ici en exposer l'ensemble des tenants et aboutissants. Je mentionnerai les travaux expérimentaux classiques de Batson qui a démontré que les personnes sont souvent motivées à aider des étrangers en détresse – et même quelques fois disposés à recevoir une décharge électrique à leur place (Batson, 1991). Bien que Cialdini ait par la suite montré que les personnes sont, par contre, moins disposées à fournir leur aide dans les situations où ils ont un moyen pour atténuer leur propre inconfort sans aider autrui (Cialdini et coll. 1987), il n'en demeure pas moins que les personnes disposées à payer un *certain* prix pour tirer autrui d'une mauvaise situation (Batson, 1998).

La coopération avec des étrangers, comme c'est d'ailleurs la norme dans les sociétés de grandes dimensions, requiert des explications différentes des mécanismes de sélection de

parentèle et d'altruisme réciproque que les biologistes ont mis en relief dans le monde animal[47]. Les humains – et peut-être quelques espèces de primates – semblent être les seuls capables de comportements altruistes *désintéressés*.

L'existence de comportements désintéressés s'accorde du reste très bien avec nos intuitions les plus fortes, en particulier quand l'agent ne connait pas et n'a aucun contact psychologique avec le bénéficiaire. Par exemple, je peux nettoyer mes contenants de lait ou de boîtes de conserve souillés afin de les mettre au recyclage bien que je n'en ai pas du tout envie. Je le fais seulement parce que j'estime devoir ne pas dégrader les conditions de vie des générations futures. Je n'en retire aucun plaisir et le fait même en bougonnant. D'ailleurs, je pourrais éprouver un plaisir réel d'avoir ainsi participé à un effort collectif de réduction des déchets, mais cela n'implique pas qu'au moment où j'ai pris la décision de nettoyer ces contenants, j'escomptais obtenir une certaine dose de satisfaction. Je l'ai peut-être fait par souci pour les générations futures, même s'il en a résulté chez moi un plaisir inattendu, lequel aurait été sans doute suffisant pour me motiver à nettoyer mes contenants.

Un partisan du réalisme égoïste comme l'économiste Gary Becker (1976) soutient que même dans ce genre de situation, on garde toujours à l'esprit une notion de notre intérêt, ne serait-ce que de manière inconsciente. Si je bougonne, c'est que je sais que mon intérêt réside dans le fait de jeter ces contenants directement aux ordures. Seulement, je ne focalise pas, suivant l'expression de Pettit, sur mon intérêt bien pesé, mais scrute invariablement de façon périphérique la ligne de conduite qu'il m'indique. Ce que le modèle focal/périphérique suggère, est

> « […] *qu'ils soient ou non conscients, ceux qui mettent en œuvre des délibérations orientées vers autrui s'autorisent un style de réflexion plus égocentrique dans leur for intérieur, aux marges de leur attention.* » (2004 : 58)

Refusant la portée explicative du modèle focal/périphérique, Pettit lui substitue celui de réel/virtuel – sans être cependant certain que ce dernier modèle est vraiment adéquat. Les agents rationnels peuvent être dans certaines situations de réels altruistes désintéressés, mais demeurer virtuellement égoïstes. On demeure virtuellement égoïste lorsqu'on est

[47] La sélection de parentèle donne lieu à des échanges coopératifs quelques fois unidirectionnels dont la prévalence et l'intensité sont fonction de la proximité génétique. Tandis que les mécanismes de réciprocité sans requérir de bagages génétiques communs ne sont possibles que dans des situations qui ont la même structure qu'un dilemme du prisonnier itératif. Or, ces formes d'altruisme ne relèvent pas, à strictement parler, de comportements désintéressés.

sensible à certaines alarmes qui s'allument quand une ligne de conduite qu'on s'apprête à adopter ou qu'on est en train d'adopter implique un coût d'opportunité pour soi-même trop important. Lorsque l'alarme retentit, les agents considèreraient leurs intérêts et les mettraient dans la balance délibérative. Cela n'implique pas pour autant qu'ils prennent une décision égoïste, mais que les considérations égoïstes permettent d'expliquer certains aspects de leur décision, par exemple, le temps utilisé pour la prendre, le manque d'empressement, la création de subtils signaux montrant que si l'on est prêt à coopérer on n'est toutefois pas disposé à nous faire exploiter, ou qu'on s'attend à un retour d'ascenseur en cas de besoin, etc.

Je ne discuterai pas davantage de la robustesse de l'hypothèse de l'altruisme désintéressé. Mon but ici est simplement de montrer qu'on peut aménager une place conceptuelle à hypothèse que les agents décident, dans certains contextes, d'agir en fonction de préférences qui peuvent être des solutions de compromis entre diverses considérations, dont certaines sont altruistes. Cela reste compatible avec l'hypothèse de la maximisation, parce que les agents peuvent avoir des préférences altruistes dont ils cherchent à maximiser la satisfaction. Seulement, on doit toujours pouvoir déduire ces préférences des choix effectifs des agents. Si un agent affirme qu'il a un souci pour autrui, mais qu'on ne peut observer chez lui aucun comportement altruiste, on doit en conclure qu'il n'a pas de telles préférences. Cela exclut aussi l'hypothèse que des agents puissent vivre des épisodes de stricte akrasie morale.

4.2.4 Le caractère déontologique/conséquentialiste des préférences

Les préférences ne portent pas seulement sur des fins, mais, dans certaines situations, sur des moyens. Les restrictions que les économistes donnent à toute théorie des préférences scientifiquement rigoureuse excluent l'idée que les agents sont non seulement mus par des considérations conséquentialistes, mais aussi déontologiques. Pourtant, cela n'est pas conceptuellement exclu par la théorie des préférences révélées.

Dans beaucoup de situations, les agents sont typiquement mus par des considérations relevant de normes autant que par des considérations relatives aux conséquences espérées (altruistes ou égoïstes), bien que leur comportement semble *prima facie* motivé exclusivement par l'une ou l'autre de ces considérations. Comme ils adhèrent à des normes

morales, légales et esthétiques qui leur interdisent d'emprunter certaines lignes de conduite, même si elles se révèlent être le seul moyen pour atteindre une fin désirée, les agents ne sont pas seulement indifférents aux conséquences de leurs actions, mais à ces actions elles-mêmes. Par exemple, la plupart des personnes ne sont pas disposées à mentir à un être cher parce que cela lui permettrait d'obtenir un dollar, même si le mensonge est sans grandes conséquences pour la personne abusée. Cela dénote chez la personne une préférence pour la sincérité ou une aversion pour le mensonge. Cette préférence ou aversion est codifiée ou enrégimentée par une norme morale qui proscrit le mensonge. Pour reprendre une suggestion faite par Joseph Heath dans *Following The Rules*, les *principes* (éthiques ou esthétiques) peuvent sans difficulté être traités comme des préférences ou aversions pour un type d'acte sans égard à ses conséquences (2008 : 71-77). Cela dit, rien ne nous permet de conclure que les agents sont exclusivement mus tantôt par des préférences déontologiques tantôt par des préférences conséquentialistes. Les préférences situées à la jonction des décisions reflètent souvent des solutions de compromis entre ces deux catégories de préférence. Si l'on reprend notre exemple du mensonge, on dira que si les personnes ne sont pas disposées à mentir pour obtenir un dollar, elles sont en revanche prêtes à mentir à un être cher si cela leur permettrait d'obtenir un million. Est-ce que dans ces deux situations les personnes sont elles exclusivement motivées par des considérations déontologiques ou conséquentialistes? Il y a tout lieu de penser que ce ne n'est ni l'une ni l'autre. La raison est qu'on est en principe capable d'identifier un point d'indifférence : il existe un montant d'agent pour lequel les personnes seront indifférentes à l'idée de mentir ou de s'en abstenir. Or, au fur et à mesure qu'on se rapproche de ce montant, les personnes prendront plus de temps pour délibérer, ce qui montre qu'elles sont en train de pondérer des considérations hétérogènes, à savoir un principe et une conséquence fort attrayante.

Cela dit, ce phénomène n'implique pas que les personnes considèrent, dans ce genre de situation, tous les montants et qu'elles leur attribuent un poids dans leur délibération. Les montants trop petits peuvent être typiquement ignorés comme le sont certaines options telles que celles qui consistent à pousser hors de son siège une vieille dame lorsqu'on cherche une place où s'asseoir dans l'autobus[48]. Les normes morales, légales et esthétiques

[48] Exemple attribué à Heath.

auxquelles les agents adhèrent n'assignent pas uniquement des poids à des options, mais leur permettent d'en ignorer certaines lorsque vient le temps de délibérer. Cela allège et simplifie grandement le processus de computation des raisons et nous rend d'autant plus efficace qu'elles (les normes) limitent les épisodes d'indécision. En effet, si vous instrumentalisez tous les rapports que vous êtes susceptibles d'avoir avec les autres (si une telle chose est possible), vous multiplierez les occasions de tricherie potentiellement praticables. Or, comme la tricherie implique des risques et un certain degré d'incertitude, qu'il est souvent difficile de mesurer en raison de son caractère stratégique, il s'ensuit que sa considération requiert une délibération plus sophistiquée.

Accepter l'idée de traiter les principes comme des préférences sur des moyens a pour effet de restreindre énormément l'étendue des cas putatifs d'akrasie morale. Les agents peuvent bien affirmer ensuite qu'ils éprouvent de la culpabilité après avoir agi à l'encontre de normes morales qu'ils chérissent, ou dire qu'ils ne se contrôlaient pas, mais cela n'exclut pas la possibilité qu'ils aient agi de manière à maximiser la satisfaction de leurs préférences. Dans le cadre de la théorie des préférences révélées, les cas putatifs de stricte akrasie qui impliquent des principes moraux, légaux ou esthétiques sont aussi chimériques que les cas putatifs d'akrasie morale qui impliquent des attitudes altruistes.

4.2.5 Quelle faiblesse de volonté pour la théorie des préférences révélées?

Le phénomène que décrit la conception (CSE) n'est pas possible si l'on estime que les choix sont les uniques indicateurs fiables des préférences. Mais cela n'implique pas que la faiblesse de volonté telle que nous semblons la voir à l'œuvre dans des comportements comme fumer après avoir pris la résolution d'arrêter, faire son rapport d'impôt systématiquement après la date limite, procrastiner lorsque vient le temps de réaliser des besognes domestiques, trop acheter avec sa carte de crédit, trop attendre avant de décider de subir un examen médical pourtant requis, etc. n'existe pas. Seulement, dans le cadre de la théorie des préférences révélées, ces comportements relèvent plus d'une inconstance dynamique des préférences que de décisions contre-préférentielles. Si j'acquiers à grands frais un appareil d'exercice que je remiserai finalement un mois plus tard, je change tout simplement de préférences. Cela n'implique pas que je ne souhaite plus être mince et en bonne forme physique, mais seulement que d'autres considérations prennent le pas, comme

vouloir regarder la télévision ou lire le dernier roman de notre auteur favori. La préférence située à la jonction des décisions va invariablement refléter une solution de compromis qui peut être à l'occasion *temporaire*.

4.3 Les écueils de la théorie

Au-delà des aspirations théoriques qu'elle incarne, la théorie des préférences révélées reste assez peu compréhensive, mais n'est pas pour autant dénuée de pertinences pratiques. À part le fait qu'elle implique une méthode qui se veut aussi simple que possible pour identifier des préférences, il est souvent judicieux de rappeler aux personnes avec qui l'on interagit qu'en dépit de leur bonne conscience et de leurs louables intentions, les choix qu'elles font trahissent en fait ce qu'elles désirent vraiment. Mais les avantages s'arrêtent ici. La théorie des préférences révélées apparait à certains théoriciens de l'économie politique beaucoup trop étriquée (Sen, 1987, 2002). Elle offre non seulement une compréhension limitée – et c'est le moins qu'on puisse dire – du comportement des agents, mais elle est aussi erronée. Examinons ici quelques-unes des critiques que je pense suffisamment sérieuses pour motiver le rejet de cette théorie.

4.3.1 Erreurs et méprises

La première critique qu'on peut lui faire est que les erreurs et les méprises sont des contre-exemples évidents. J'achète un kilo de café moka-java à l'épicerie et me rends compte à la maison que je me suis trompé, et je pensais avoir acheté du café nigérien. Je me dis tant pis et je le bois quand même. Mon choix ne reflète en rien ma préférence pour une essence de café. Un partisan de la théorie des préférences révélées pourrait soutenir que mon choix reflète tout de même une préférence pour le café par rapport aux autres boissons. L'argument est mauvais pour deux raisons. Je peux préférer un thé rouge au moka-java, auquel cas j'aurais choisi ce thé si je n'avais eu le choix qu'entre ces deux alternatives. L'autre raison est que ma préférence pour le café nigérien n'est tout simplement pas révélée par mon choix et qu'on a besoin d'une autre source d'information pour pouvoir l'identifier. Les erreurs sont, du reste, fréquentes et ne peuvent pas être considérées comme des anomalies à la frontière de la théorie, et qu'on peut tout simplement ignorer. Les habitudes fortement ancrées et la prise en charge de nos lignes de conduite par des systèmes cognitifs automatiques qui font l'économie du maintien de l'attention nous font à l'occasion faire des

choix contre-préférentiels. Je peux par exemple prendre la décision de téléphoner à un ami et composer sans m'en rendre compte le numéro d'un autre ami que j'ai l'habitude d'appeler régulièrement.

4.3.2 La fausse indifférence

La théorie des préférences révélées stipule que si les agents ne sont pas capables de choisir entre deux alternatives, alors ils y sont indifférents. Ceci est plutôt contre-intuitif. Dans certains contextes de choix, l'indécision ne reflète pas une indifférence à l'égard des alternatives, mais plutôt une difficulté à traiter les informations qui nous permettraient de trancher. Nous pouvons suspendre notre décision si nous avons à choisir entre deux emplois qui représentent beaucoup pour nous, ou entre sa femme et sa maitresse, ou entre deux maisons qui comportent des avantages et des inconvénients différents et difficilement comparables. Mais cela n'implique pas que nous sommes indifférents au choix de ces options.

4.3.3 Automatisme douloureux

Il semble que les agents s'adonnent à l'occasion à des activités déplaisantes qu'ils ne veulent pourtant pas faire. Par exemple, William James mentionnait que les personnes qui ont mal à une dent ont tendance à la tâter continuellement du bout de la langue, ce qui produit à chaque fois une douleur vive. Pourtant, on ne peut pas en inférer que ces personnes ont une préférence pour la douleur. Un partisan de la théorie des préférences révélées pourrait arguer que le tripotage douloureux de la dent est compensé par le fait d'acquérir de l'information afin de savoir si elle fait toujours mal. Il s'agit d'une réponse bizarre parce que les maux de dents ne disparaissent pas sans interventions médicales, et aussi parce que quand une partie de notre bouche est la source d'une douleur lancinante, on peut être à peu près certain que la tripoter accentuera la douleur *sans* même être en train de le faire actuellement (Hastie & Dawes, 2010). Une réponse insatisfaisante serait de dire que les agents ont des goûts qui ne se discutent pas (*de gustibus non disputandum*). Nous n'avons pas besoin d'être masochistes pour nous adonner automatiquement à ce genre d'activité.

4.3.4 Exclusion des indicateurs hétérogènes

Cela m'amène à la seconde critique qui porte sur les indices révélateurs de préférences. À côté des choix, nous disposons de toute une batterie d'indicateurs de nature hétérogène pour connaître les préférences des agents. Les descriptions que les agents font de leur propre comportement sont sans doute la source qui offre le degré le plus élevé de précision. C'est pourquoi on est enclin à questionner les agents lorsque leur comportement nous semble quelque peu énigmatique. Or, la possibilité du mensonge, de la duperie de soi ou tout simplement du manque de connaissance de soi-même montre les limites des comptes rendus en première personne, d'où l'utilité d'avoir plusieurs indicateurs. Les expressions faciales et l'orientation du regard sont également des indicateurs précieux et nous permettent dans certaines circonstances d'inférer avec relativement peu d'ambiguïtés des préférences. En fait, nous nous servons bien souvent – mais pas toujours – de ces indicateurs en paire et de manière séquentielle. Nous observons d'abord l'orientation du regard de l'agent, nous savons alors sur quoi il fixe son attention. Nous observons ensuite le rictus de l'agent pour déterminer s'il éprouve du plaisir ou du désagrément. Nous pouvons alors inférer que l'agent a une préférence ou une aversion (qui est en fait une préférence pour éviter quelque chose) pour l'objet de son regard. On peut raffiner ces indicateurs et mesurer en plus le temps de fixation de l'attention et de vitesse de formation du rictus. Cela nous permet d'obtenir un ordonnancement plus détaillé des préférences. D'ailleurs, ces règles inférentielles sont si importantes pour la vie sociale qu'elles sont implémentées de manière innée dans le cerveau des enfants. Elles font partie de ce que les psychologues appellent les théories de l'esprit, soit un dispositif cognitif qui permet aux organismes de faire des hypothèses sur les états mentaux d'autrui et d'eux-mêmes. Une avarie ou un défaut de construction rend l'identification des préférences particulièrement laborieuse. Les enfants autistes, qui sont réputés avoir des difficultés à attribuer des états mentaux, ne réussissent pas le test de l'orientation du regard et de la forme du rictus. Lorsqu'on leur présente une scène où l'on voit un visage du personnage souriant qui fixe son regard sur un objet figurant dans un panier de biens disponibles et qu'on leur demande ensuite de déterminer ce que préfère le personnage, ils semblent donner dans la plupart des cas des

réponses aléatoires (Baron-Cohen & Frith, 1985 ; Baron-Cohen, 1999) [49], alors que des enfants de 3 ou 4 ans y parviennent sans problèmes. Le plus étrange est que les enfants autistes sont capables d'identifier les objets que les agents regardent et leurs expressions faciales. Seulement, ils n'en infèrent pas spontanément des préférences. Ils doivent apprendre ces règles et faire initialement des inférences conscientes. Cela dit, ces règles ont également leur limite. Elles ne sont applicables que dans des situations où les biens, événements ou personnes susceptibles de faire l'objet d'une préférence ou aversion, sont présentes sous forme de stimuli visuels immédiats. Et la tromperie est toujours possible. D'où l'idée que nous devions disposer d'indicateurs supplémentaires au cas où les prévisions que nous faisons à partir des préférences hypothétiquement inférées se révéleraient inadéquates. Aussi, si nous nous contentons des choix effectifs des agents pour déterminer leurs préférences, nous nous condamnons inévitablement à une cécité mentale qui nous rapproche en quelque sorte de la condition des autistes!

4.3.5 Ambigüités

Une autre critique qu'on peut adresser à cette théorie est qu'elle présente un degré trop élevé d'*ambiguïté* dans l'identification des préférences. Je peux être assis à table avec ma femme alors que nous nous apprêtons à manger un dessert. Il y a une part de gâteau au fromage et une part de tarte au citron. Je choisis la tarte. Qu'est-ce que je préfère? Il n'est pas évident de répondre à la question en nous basant seulement sur l'indice du choix effectif. Je peux préférer la tarte au citron au gâteau au fromage, comme je peux préférer faire plaisir à ma femme en lui lassant ce que je pense qu'elle préfère[50]. La dernière possibilité reste tout à fait compatible avec le fait que j'ai une préférence pour le gâteau au fromage. Mon choix reflète-t-il une préférence de nature altruiste ou une préférence alimentaire? Il faudrait examiner d'autres choix qui impliquent des desserts sucrés pour

[49] Les auteurs mettent en relief pour la première fois les capacités d'attribution d'états mentaux des autistes en les contrastant avec ceux des enfants normaux et des enfants trisomiques, notamment sur l'attribution de croyances fausses à partir du test Sally-Anne.

[50] En fait, ces deux éventualités ne sont pas incompatibles. Je peux avoir ces deux types de préférences. Mais si l'on veut expliquer pourquoi j'ai choisi la tarte, on doit identifier *la* raison qui m'a motivée à faire ce choix, soit la raison qui avait le plus de poids dans ma délibération ou celle qui ferait hypothétiquement la différence dans des situations du même type. Certaines conséquences des décisions des agents peuvent être prévues et même voulues par eux sans qu'on puisse en inférer que ces conséquences constituent la raison qui explique pourquoi les agents ont pris telle ou telle décision. Je peux, par exemple, savoir que, si je me rends au salon de coiffure, j'aurai droit aux derniers potins sur la vie des vedettes hollywoodiennes et souhaiter que cela arrive. Mais cela n'explique pas le choix de me rendre au salon. Je peux m'y rendre parce que veux que mes cheveux soient coupés.

pouvoir identifier ma préférence. Or, il persiste toujours un degré d'ambiguïté assez élevé pour que cette théorie nous apparaisse insatisfaisante. Par exemple, je peux, une semaine plus tard, avoir une fringale. J'ouvre alors mon réfrigérateur et aperçois encore une part de gâteau au fromage et trois parts de tarte au citron. Je choisis encore la tarte. Qu'est-ce que je préfère? Si l'on se base strictement sur l'observation de mes choix, *il semble* que je préfère la tarte. Mais si vous me le demandez, je répondrai le gâteau. Je vous expliquerai mon choix en disant, la première fois, que je savais que ma femme préfère également le gâteau au fromage et que je voulais lui laisser. La seconde fois, je savais que la tarte était beaucoup plus fraîche que le gâteau et c'est pour cela que je l'ai choisie. Sur la base exclusive des choix que j'ai effectivement faits, peut-on établir que j'aie une préférence alimentaire pour les tartes au citron, les gâteaux au fromage, les tartes au citron fraîches, les gâteaux au fromage frais, les tartes au citron dans un contexte social, les tartes au citron dans un contexte où ma femme est présente? Même si nous ajoutons à la liste d'indices un plus grand nombre de choix dans l'espoir de réduire l'ambiguïté, nous multiplions ce faisant les contextes de décision, donc les traits susceptibles de rentrer dans la description des préférences. D'ailleurs, rien ne nous empêche de stipuler que dans chaque contexte l'agent ait pris une décision en fonction de préférences qui ont tout simplement changé. En effet, rien, dans les choix effectifs de l'agent, ne nous permet de supposer que leurs préférences sont stables dans le temps.

Le problème de l'ambiguïté est souvent présent même lorsqu'on utilise une batterie d'indices qui ne reposent pas exclusivement sur les choix. Une personne peut rougir si on lui offre une part de gâteau. Si elle accepte d'en prendre, on peut inférer qu'elle a une préférence marquée, mais qu'elle est quelque peu honteuse pour ce type de gâterie ou qu'elle a une préférence inavouée pour la personne qui lui offre, etc. Mais à la différence des préférences révélées strictement dans les choix, on peut toujours éliminer des préférences et converger vers une solution en multipliant les indices. Pourquoi? Parce que ces indices sont non seulement hétérogènes, mais permettent d'éliminer des préférences. On ne rougit pas, par exemple, à propos de n'importe quelles préférences. Seules les préférences incontinentes, sexuelles ou amoureuses susceptibles d'une désapprobation sociale subtile – ou allant à l'encontre des conventions – produisent un afflux de sang au visage lorsqu'elles sont remarquées par autrui. Cela permet de faire un premier filtrage. On

éliminera, par exemple, des préférences comme désirer être vue en train de manger du gâteau. Ce premier est quelquefois suffisant pour qu'on adopte la bonne attitude à l'égard de la personne, et quelques fois non. Nous chercherons alors à réviser ou raffiner l'hypothèse en identifiant d'autres indices. Si la personne semble particulièrement maladroite et peu spontanée lorsque nous lui adressons la parole, nous pourrons en inférer qu'elle a un béguin pour nous et non pour les gâteaux. Aussi, exhiber des comportements maladroits et non spontanés ne relève pas d'un choix. C'est plus quelque chose qui nous arrive que quelque chose que nous faisons. Mais rien n'empêche que nos indices supplémentaires soient des choix. Les comportements mimétiques sont souvent de bons indicateurs d'une préférence sexuelle ou amoureuse. Le problème de l'indétermination devient impossible à résoudre lorsque nous nous servons exclusivement des choix effectifs des agents pour identifier leurs préférences.

4.3.6 Indétermination

Le problème de l'ambiguïté dans la théorie des préférences révélées a également été mis en relief par Davidson, mais d'une façon plus radicale. En fait, Davidson soutient que la théorie des préférences révélées souffre d'un problème plus grave d'*indétermination*. Selon Davidson, pour connaître les préférences des agents, on doit tenir pour acquis qu'ils ont certains désirs et certaines croyances. Le problème est qu'on ne peut faire les deux en même temps. C'est analogue à une équation à deux variables telles que : $x + y = 0$. Il existe plus d'une solution possible. Dans le cas des préférences, je peux, apercevant Julien courir poursuivit par un ours, expliquer son comportement en stipulant qu'il pense qu'un ours le poursuit et qu'il désire manifestement sauver sa peau, ou en stipulant qu'il pense qu'un agent du fisc le poursuive pour fraude et qu'il désire manifestement échapper au fisc, etc. Lorsque nous attribuons des préférences aux agents, nous le faisons à partir de toute une série d'hypothèses auxiliaires relevant de variables contextuelles (ex. : Julien a sans doute vu l'ours) et non contextuelles (ex. : Julien a une très bonne vision) et du principe de charité qui vaut que les agents ont la plupart du temps des croyances et désirs rationnels. Or, toutes ces hypothèses sont exclues lorsqu'on décide de se baser uniquement sur les choix effectifs des agents.

4.3.7 Empirisme naïf

D'ailleurs, les problèmes d'ambiguïté et d'indétermination sont tributaires d'une sorte d'empirisme naïf qui, se contentant des choix effectifs comme base d'inférence des préférences, ne nous permet pas de faire des prédictions non triviales. Si un agent choisit *a*, alors il préfère faire ou obtenir *a*. Si nous souhaitons prédire ce que choisira un agent, tout ce que nous pouvons dire c'est que s'il préfère *a*, alors il choisira *a*! Or, l'utilité d'une théorie des préférences est qu'elle est censée nous aider à prédire ce que les agents feront, et ce, d'une manière non triviale. Une théorie non triviale repose sur toute une série d'hypothèses sur ce que les agents feraient dans telle ou telle situation ou sur ce qu'ils auraient fait s'ils y étaient plongés (hypothèses contrefactuelles). En testant les premières, nous pouvons lever l'ambiguïté des choix et arriver à raffiner nos attributions de préférence. En théorie de la décision, nous gardons toujours à l'esprit une situation hypothétique idéale dans laquelle l'agent satisfait sa préférence. Si j'ai une préférence pour les gâteaux au fromage, alors il existe une situation idéale où j'ai faim et où mon acte n'a aucun impact sur les autres ni sur moi-même (si ce n'est le plaisir qu'il me confère, etc.) et où je décide d'opter pour le gâteau et non pour la tarte au citron. Lorsque les économistes font des prédictions reposant sur les préférences des agents (consommateurs, électeurs, etc.), ils s'appuient en fait sur toutes sortes d'hypothèses auxiliaires, factuelles et contrefactuelles, pour identifier ces choix. Comme ces hypothèses figurent implicitement dans leurs calculs, les théoriciens ont tendance à croire qu'ils peuvent en faire l'économie, ce qui confère une sorte de crédit immérité à la théorie des préférences révélées.

Cela dit, l'empirisme naïf de la théorie des préférences révélées fait d'elle non seulement une théorie peu informative, mais elle repose également sur une position épistémologique moribonde, à savoir que nos meilleures théories doivent être surdéterminées par l'expérience empirique. Or, depuis la publication du célèbre article de Quine (1951), *Les deux dogmes de l'empirisme*, il y a un relatif consensus parmi les philosophes sur le fait que nos meilleures théories soient résolument sous-déterminées par l'expérience – ce qui n'est pas, comme Quine le souligne, incompatible avec l'idée que seule l'expérience détermine, en fin de compte, nos meilleures théories.

Je ne crois pas que la théorie des préférences révélées ait beaucoup d'ascendant sur la communauté philosophique. Elle est d'ailleurs beaucoup critiquée en micro-économie et

surtout dans le domaine de l'économie comportementale. Il s'agit d'une version radicale et par trop simplifiée de la conception platonicienne de la motivation. Elle ne permet ni d'expliquer pourquoi la stricte akrasie est possible ni de comprendre la faiblesse de volonté en général. Pour ce faire, une théorie de la motivation qui décrit plus finement le processus de décision nous est nécessaire.

4.3.8 Les effets de comparaison

Les problèmes décisionnels non triviaux auxquels font face les agents rationnels dans la vie de tous les jours requièrent l'examen de toutes sortes de considérations *avant* qu'ils prennent effectivement une décision. Cela peut prendre quelques millisecondes à plusieurs heures – suivant la nature des choix à faire – et peut être très simple ou très complexe. Aussi, il est méthodologiquement étrange de ne pas en considérer le détail pour spécifier la nature des décisions et des préférences des agents et isoler le problème akratique.

D'ailleurs, bon nombre de résultats expérimentaux et de leurs interprétations mettent en relief le fait que les agents établissent ou construisent leur préférence *en comparant* des alternatives et ne consultent pas une sorte de menu interne (ordonnancement des alternatives) qui leur indique clairement et sans ambigüité ce qu'ils doivent faire.

Par exemple, le comportement des consommateurs qui choisissent x dans la suite $[x, y]$ et y dans la suite $[x, y, z]$ peut sembler étrange. En fait, les consommateurs peuvent choisir majoritairement x dans la première série où il est difficile de comparer les options de cette série. En revanche, si y est facilement comparable à z et exemplifie des propriétés relationnelles plus appréciables, alors les consommateurs

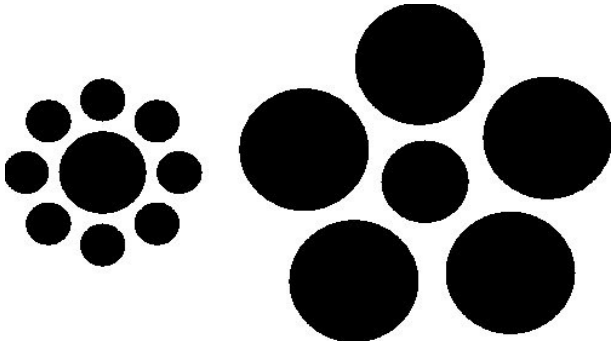

Figure 1

opteront en majorité pour *y* si la seconde suite leur est présentée[51]. Dan Ariely (2008 : 1-21) illustre ce phénomène à l'aide d'une publicité sur laquelle il est tombé alors qu'il feuilletait une revue scientifique. La publicité offre 3 modalités d'abonnement à la revue *The Economist* : la première offrait un abonnement annuel à une version web au coût de 59 $, la seconde offrait une version papier seulement pour 125 $, tandis que la dernière offrait une version web et papier pour 125 $. Ariely a cru d'abord que l'offre comportait une erreur : qui voudrait de la seconde option? En sondant des étudiants du MIT susceptibles de vouloir un abonnement à *The Economist*, Ariely a établi que la seconde option joue le rôle de faire-valoir (*decoy*) pour la troisième. Les résultats qu'il a obtenus en sondant deux groupes de 100 étudiants auxquels on présentait l'offre telle quelle et l'offre sans la présence du faire-valoir sont sans équivoques[52].

Puisque la présence d'un faire-valoir modifie de manière significative les préférences des agents, la meilleure explication est que ceux-ci *établissent* leur préférence *en comparant* les options plutôt qu'en *consultant* une sorte menu interne leur indiquant leur valeur respective. Ariely compare l'effet de comparaison de nos préférences avec l'illusion bien connue des cercles. Lorsqu'entouré de petits cercles, un cercle semble plus gros que lorsqu'il est entouré de cercles plus grands. Bien que les propriétés intrinsèques de ces deux cercles soient les mêmes, leurs propriétés relationnelles produisent en nous l'illusion qu'ils sont de tailles différentes (figure 1).

4.4 Conclusion

Une conception générale plus réaliste et riche de la motivation a besoin de décrire ce qui se passe en amont des décisions avec des clefs variées, et pas uniquement les choix des agents. Adhérer à l'hypothèse de la maximisation implique que si nous devons voir les préférences seulement à la surface des décisions, alors nous devons exclure l'existence de cas de stricte akrasie tels que décrits par (CSE). Or, voir les préférences à la surface des décisions pose

[51] En économie comportementale, ce phénomène est décrit comme un cas de dominance asymétrique qui viole l'axiome de l'indépendance des options.

[52] Pour le premier groupe, la première option avait la préférence de 16 personnes, la seconde aucune, tandis que la troisième était préférée par 84 personnes. Pour le second groupe, la première option était préférée par 68 personnes, tandis que la troisième ne l'était que par 32 personnes.

tant de problèmes que cela ne peut pas constituer une bonne raison d'être sceptique à l'égard de la stricte akrasie.

Une position sceptique (ou non sceptique) plus robuste doit reposer sur une conception beaucoup plus riche de la motivation et des processus décisionnels. S'il existe un problème de stricte akrasie, celui-ci devrait hypothétiquement être situé en amont de la décision quelque part dans le processus de comparaison des alternatives. Or, la théorie des préférences révélées rend le processus complètement opaque, et il devient par le fait même impossible d'identifier le lieu, l'étape ou le mécanisme qui serait en cause dans les cas de stricte akrasie. Une théorie robuste devrait montrer comment la stricte akrasie est, à une étape particulière du processus, possible ou impossible.

Bien qu'on ne trouve pas de descriptions détaillées du processus décisionnel dans la littérature philosophique qui traite de la question de la faiblesse de volonté, il y a un certain consensus sur la nature de l'étape du processus qui poserait problème pour l'agent strictement akratique. Le prochain chapitre sera consacré à la description du processus et l'identification hypothétique du problème de la stricte akrasie. La description exploitera certaines idées classiques comme la maximisation et l'idée que les agents rationnels procèdent à une délibération pour identifier le meilleur choix, mais aussi l'hypothèse plus récente suivant laquelle deux types de processus en bonne partie distincts peuvent prendre en charge nos décisions.

5
Le problème akratique et la formation du jugement et des préférences

Decision making is a specific executive task.

–Peter F. Drucker

A good image of what we mean by decision making is of a person pausing at fork in the road, and then choosing on path – to reach a desired goal or avoid an unpleasant outcome.

–Robyn Dawes et Reid Hastie

5.1 Le processus décisionnel putatif de l'agent akratique

Comme nous l'avons vu au chapitre précédent, il est incorrect de nier l'existence de la stricte akrasie en se basant uniquement sur une théorie des préférences révélées. Lire les préférences à la surface des décisions comporte beaucoup trop de problèmes théoriques pour que cela justifie une position sceptique robuste. On doit nécessairement être en mesure de décrire les processus qui amènent les agents à décider (rationnellement ou non) en faveur d'une ligne de conduite ou d'une autre. S'il y a un débat à avoir sur la nature et l'existence de la faiblesse de volonté, on doit montrer à quel niveau elle est censée se produire dans le processus décisionnel. Spécifier ce ou ces niveaux permet non seulement de camper correctement le débat, mais également d'éviter certaines sources de confusion. Décrire les processus décisionnels putatifs de l'agent rationnel permet aussi d'identifier plus précisément les conditions sous lesquelles la conception Standard pourrait être vraie et ses cas limites.

La description des processus décisionnels des agents présentée dans ce chapitre doit figurer en toile de fond du débat philosophique sur l'existence qui a court depuis les 50 dernières années. Il est bien important de la comprendre, parce que j'aborderai le débat comme tel dans le prochain chapitre.

Je ne prétends pas que la description des processus décisionnels que j'exposerai ici, ait été dans l'ensemble, défendue explicitement par un théoricien qui aurait pris part au débat.

Mais on retrouve plusieurs de ces éléments explicitement ou implicitement avancés ou défendus dans la littérature philosophique, économique et psychologique. Aussi, certains de ces éléments, sans avoir été d'une quelconque manière avancés ou défendus, *auraient dû* l'être (ex. : l'existence de processus automatiques et de planification irréaliste) pour éviter certaines confusions ou du moins préciser certaines limites de la conception Standard.

5.1.1 Problème décisionnel et modèle du « *double-processus* »

Nous prenons régulièrement des décisions allant à l'encontre de notre meilleur jugement. Par exemple, j'avais prévu me rendre à l'aéroport avec mon automobile pour aller chercher un ami, seulement je n'ai pas pris la bonne sortie et ai emprunté plutôt le chemin que j'emprunte régulièrement pour me rendre à mon lieu de travail. Ai-je pour autant vécu un épisode de faiblesse de stricte akrasie? Bien que des décisions de la sorte satisfassent la conception Standard, elles ne semblent pas relever intuitivement de la faiblesse de volonté.

Chaque jour, nous accomplissons des milliers d'actions : nous déambulons vers des destinations particulières, nous respirons, nous nous grattons, nous bougeons la tête, nous levons le bras droit, nous nous assoyons, nous sortons les poubelles ou nous faisons le ménage, nous lisons, nous nous couchons, nous parlons, etc. Parmi ces actions, certaines sont faites de manière plus ou moins consciente, alors que d'autres requièrent toute notre attention. Dans le cas des premières, la quantité de décisions conscientes que nous prenons pour les accomplir est très restreinte. Je peux, par exemple, décider de marcher jusqu'à mon lieu de travail au lieu de prendre ma voiture. Cela n'implique pas que je devrai prendre une multitude de décisions conscientes concernant la quantité de pas que je ferai, le rythme de mes pas, le choix de la jambe à mettre ma marche en branle (devrais-je commencer par la droite ou par la gauche?), le choix de la portée de mes pas, etc.

Tous ces aspects sont en quelque sorte déterminés par un *système automatique* de prise de décision. Ce système prend en charge la plupart des tâches que nous accomplissons lorsque nous nous adonnons à des activités particulières, comme la marche. Ce système comporte des avantages évidents : il est moins énergivore (il est moins fatigant), très discret (il n'encombre pas nos esprits de toutes sortes de pensées), ne nécessite pas beaucoup de concentration (il peut piloter un nombre très important de tâches différentes simultanément) et est très rapide. Ce système peut être inné – comme dans le cas des mécanismes-réflexes –

ou acquis – comme dans le cas des habitudes bien implantées dans la personne. Dès lors que nous apprenons à marcher ou à attacher nos lacets, nous n'avons plus besoin de prendre toute une série de décisions ponctuant chacune des étapes de ces activités. Par exemple, dans le cas des lacets, à présent que je sais comment procéder, je n'ai plus à me demander si je dois commencer par le bout droit du lacet ou par celui de gauche, ce que je dois en faire, et, une fois terminé avec le premier, ce que je dois faire avec l'autre, etc. À la rigueur, tout ce dont je dois me demander est si je devrais faire une boucle simple ou double pour plus de solidité, parce que je compte aller faire une randonnée en montagne.

Aussi, la plupart des choses que nous faisons dans une journée sont pilotées par un système automatique. Mais il arrive des situations ou ce système n'est pas très efficace et, si nous continuons à nous fier à lui, nous courons tout droit à la catastrophe! Les parents apprennent à leurs enfants à toujours regarder de chaque côté de la rue avant de traverser parce qu'ils savent que les enfants sont très impulsifs et qu'ils ont du mal à engager un processus décisionnel. Or, leur impulsivité les pousse souvent à prendre des risques importants – risques qu'ils pourraient pourtant éviter facilement. Aussi, on inculque aux enfants à *inhiber* leur système automatique au moyen d'une sorte de dispositif de contrôle de soi (dont le siège est situé en bonne partie dans le cortex préfrontal) et à *activer* leur *système contrôlé* de prise de décision. Le système manuel prend en charge le processus de délibération consciente. Une fois la décision prise à ce niveau, le reste du travail est fait par le système automatique, jusqu'à ce qu'un nouveau problème décisionnel surgisse.

Le modèle du double-processus (*dual-process*) est un objet d'investigation poussée dans les sciences cognitives actuelles. Il remonte en fait à William James (1890) qui distinguait la pensée associative du vrai raisonnement. Il estimait que le contenu représentationnel de pensée associative est déterminé par les expériences passées, alors que celui du raisonnement véritable porterait sur des situations originales ou sans précédents. La description du fonctionnement des deux processus computationnels s'est raffinée notamment depuis la publication des travaux de Kahnemann (1999, 2004) sur les heuristiques du jugement . Le psychologue Keith Stanovich (2004) est sans doute un des plus fervents promoteurs de ce modèle. En dehors de celles mentionnées plus haut, les propriétés des systèmes qui pilotent, gèrent et monitorent les deux processus peuvent être

très abstraites[53], mais nous sommes capables, la plupart du temps, de distinguer les moments où nos comportements sont pilotés par nos systèmes automatiques et contrôlés. Nous savons intuitivement qu'il y a une différence qualitative importante entre la manière dont nos activités sont prises en charge par ces deux systèmes. Il y a une différence entre prendre la décision de mettre un pied devant l'autre, de faire une inférence, de classifier des objets, de les dénombrer, de faire une hypothèse, etc., et faire ces choses de manière automatique sans y penser. D'ailleurs, la confusion des processus de pilotage des activités donne quelques fois lieu à des quiproquos. Les individus qui ont sauvé des personnes au péril de leur vie ne comprennent souvent pas pourquoi on devrait les considérer comme des héros. Ils soutiennent qu'ils étaient dans une sorte d'état second où ils ne mesuraient pas les risques avec acuité et ne semblaient même pas « s'appartenir » tout à fait. Aussi, ils estiment que n'importe qui aurait fait la même chose parce que les circonstances imposaient le comportement qu'ils ont adopté. On attribue spontanément cette réaction à de la vraie ou fausse humilité, alors qu'une autre hypothèse est sans doute plus probable. Ces héros ont accompli des gestes étonnants parce que leur ligne de conduite n'a pas été pilotée par un système de planification intentionnelle, mais par le système automatique. De manière analogue, c'est ce qui se passe lorsque nous faisons face, sans préparation, à une situation d'urgence comme un incendie ou lorsqu'une voiture fonce sur nous.

Cela dit, si ces mécanismes fondamentaux se situent typiquement en deçà de la conscience, cela n'implique pas que les agents deviennent des zombies ou des pantins mus par des

[53] Dans ce modèle, le système automatique se distingue notamment par le fait qu'il est capable de traiter en parallèle les informations nécessaires à l'accomplissement d'une tâche complexe, ce qui explique sa rapidité. Tandis que le système contrôlé traite les informations de manière sérielle, ce qui explique sa lenteur relative. Délibérer est un processus conscient de traitement de l'information en série. Nous ne pouvons évaluer qu'une seule option, ou pièce d'information, à la fois. Le processus est analogue à celui qui consiste à lire des phrases. Nous ne pouvons fixer notre attention que sur un seul mot à la fois; et pour comprendre le texte, nous devons passer en revue les mots dans l'ordre prescrit par celui-ci. En revanche, lire les mots contenus dans une phrase fait davantage appel à un traitement en parallèle des informations. Nous traitons les mots écrits comme des touts. À part la première et la dernière lettre, nous pouvons changer l'ordre de toutes les autres sans en altérer significativement la compréhension : « Puvoez-vuos cpromenrde ce que vuos lesiz en ce mmeont? » Le truc consiste à lire la phrase rapidement de manière à nous assurer que notre système manuel n'essaye pas de prendre le commandement du traitement de l'information.

Les chercheurs ont identifié beaucoup de propriétés distinctives qu'ont ces systèmes et la nature des tâches qu'ils permettent d'accomplir, et cela, en particulier dans le domaine de la cognition – et moins dans le domaine de la motricité. Les propriétés du système automatique sont holistiques, automatiques, associatives, cognitivement économiques, relativement rapides, acquises biologiquement, par exposition répétée et à la faveur d'expériences personnelles. Les propriétés du système contrôlé sont analytiques, contrôlées, basées sur des règles, cognitivement exigeantes parce qu'elles requièrent de fixer son attention, relativement lentes, acquises culturellement et de manière plus formelles. Les tâches auxquelles le système automatique est adapté sont hautement contextuelles, personnalisées, de nature conversationnelle et sociale, tandis que les tâches auxquelles le système contrôlé est plus adapté sont non contextuelles, dépersonnalisées et non sociales. Pour un résumé des travaux sur le sujet, voir Stanovich (1999 : 144-148).

causes « extérieures » à leur esprit. En fait, nous pouvons faire des choses qui sont gérées par nos systèmes automatiques et contrôlés en même temps. Je peux délibérer pour savoir si je vais me faire un repas gourmet ou aller au restaurant tout en marchant et en mâchant de la gomme[54].

Qu'est-ce que tout cela nous apprend sur la faiblesse de volonté? Nous ne pouvons pas considérer que les activités qui sont entièrement pilotées par le système automatique peuvent être akratiques. Je peux délibérer et arriver à la conclusion qu'il est préférable que je me rende chez mon ami Steve. Je peux emprunter le chemin de sa maison et déambuler

[54] On peut critiquer l'hypothèse que le système automatique *peut prendre des décisions*. On réserve habituellement le terme « décision » à des événements mentaux conscients. Mais nos intuitions ne sont pas toujours cohérentes à ce sujet. Nous critiquons volontiers une personne qui a pris l'habitude de faire des remarques désobligeantes même si elle ne s'en aperçoit pas. Nous estimons qu'il s'agit en quelque sorte de choix pour lesquels elle a une responsabilité même s'ils ne sont pas délibérés.

D'un point de vue plus technique, certains pourraient tout de même contester le fait de considérer les choix pilotés par le système automatique comme d'authentiques choix parce que les comportements qu'il engendre ne sont pas des actions. Or, s'il ne s'agit pas d'actions, mais de simples mouvements du corps, il ne peut y avoir de décisions intentionnelles en amont. Ce genre de critique atteint sa cible seulement si l'on adopte le critère causal de l'action.

Le *critère causal* stipule que ce qui distingue les actions des seuls comportements – ce que je fais de ce qui m'arrive – est que ces premières sont causées par des croyances et des désirs, alors que ce n'est pas le cas des secondes (Davidson, 1971 ; Goldman, 1970). Aussi, quand nos activités sont exclusivement gérées par le système automatique, les croyances et les désirs semblent n'y jouer aucun rôle. Or, beaucoup de ces activités sont indubitablement des actions et non de simples comportements. Il s'ensuit que le système automatique ne peut piloter (gérer, monitorer, etc.) à lui seul des actions, lesquelles doivent être pilotées par des systèmes intégrés. Cet argument n'est pas convaincant. Il est vrai que les activités que les psychologues estiment être pilotées ou prises en charge exclusivement par le système automatique sont souvent des actions. Le problème vient peut-être du critère causal. Indépendamment du fait qu'on adhère ou non au modèle du double processus, le critère causal semble trop restrictif. Il ne permet pas de catégoriser des structures détaillées du comportement comme des actions. Je peux dire que ma déambulation dans un corridor est une action parce qu'elle est causée par mon désir de me rendre à mon bureau et la croyance que c'est le moyen le plus approprié de m'y rendre. Mais que dire du détail de ma ligne de conduite? Est-ce que chacun de mes pas peut être expliqué par des attitudes propositionnelles? Je peux le faire, mais l'explication aura l'air soit particulièrement artificielle, soit *ad hoc*. Je peux dire que le troisième pas fut causé par le désir d'avancer et la croyance que, étant donnée la répartition de mon poids, il est approprié que j'avance cette jambe, etc. Mais je devrai inventer quelque chose de plus alambiqué pour expliquer le fait que j'ai, sans m'en rendre compte, posé le pied sur exactement une dalle à la fois jusqu'à mon bureau.

On peut alléguer que l'on ne peut pas connaître le détail de chacun de nos comportements, mais que cela n'implique pas qu'ils ne peuvent être des actions. Il doit exister une description vraie sous laquelle mon comportement est intentionnel, et si mon comportement est intentionnel sous au moins une description, alors il est causé par un désir et par une croyance (Davidson, *Ibid.* : 71.). En fait, je peux décrire chacun de mes pas en disant que je me rends à mon bureau. Si cette suggestion est correcte, alors rien n'empêche de considérer que les comportements pris en charge par le système automatique ne peuvent pas être des actions. On peut toujours redécrire une série complexe de comportements pilotés par le système automatique comme des phases menant la réalisation d'un but.

Cela dit, le critère causal de l'action n'est pas le seul ni le meilleur. Le *critère de la guidance* proposé par Frankfurt (qui s'inspire beaucoup de propositions faites par Nagel (1977 : 271)) est plus approprié et est aussi tout à fait compatible avec l'idée que le système automatique peut piloter seul des actions. Pour Frankfurt,

> « *Les comportements sont intentionnels (purposive) quand leur cours est sujet à des ajustements lesquels compensent les effets des forces qui autrement interféreraient avec le cours du comportement, et que l'occurrence de ces ajustements ne sont pas explicables par les états de choses qui les provoquent. Le comportement est dans ce cas sous la guidance d'un mécanisme causal indépendant, et dont l'empressement (readiness) à créer des ajustements tend à assurer que le comportement est accompli.* » (1978 : 47-48.).

Ce n'est pas parce que le système automatique pilote des activités sans recourir à l'attention consciente que ces activités ne peuvent être considérées comme des actions. Tout ce qui est requis est qu'il existe des dispositifs qui alertent l'agent lorsque quelque chose ne tourne pas rond et qui lui permettent d'opérer des ajustements en cas de besoin. Pour utiliser une métaphore, je dirais que ce dispositif est composé d'un service de monitorage et d'un service d'intervention musclée.

sans porter attention à ce que je fais et me retrouver chez mon ami Philippe. Toutes choses considérées, il était préférable que je me rende chez Steve que de me rendre chez Philippe. J'ai donc accompli une action contre-préférentielle. Mais il ne s'agit pas d'un épisode de faiblesse de volonté. Si je m'étais rendu compte en chemin que je me rendais chez Philippe, alors le diagnostic aurait pu être différent. En fait, mon dispositif d'ajustement s'est trop occupé de certains détails de ma conduite (équilibre, alternance des pas, enjambées adéquates, obstacles évités, etc.) et pas assez d'autres (la rue de Steve est parsemée de chênes matures, alors que celle de Philippe est nue). Aussi, les alarmes qui auraient dû sonner ne l'ont pas fait.

Bien que ce ne soit pas toujours pour des raisons détaillées, voire même claires, les philosophes contemporains excluent d'emblée les actes accomplis à la faveur d'un manque de contrôle intentionnel, et sans être précédés de réflexion, de la catégorie des actes akratiques, tandis qu'Aristote les confondait. Le psychologue Keith Stanovich – qui a beaucoup documenté les apports distincts des deux systèmes – a mis en relief les difficultés qu'éprouvent certains individus à inhiber le système automatique (Stanovich, 2004 ; Stanovich et coll., 2008). Selon lui, ce qu'échouent à montrer les tests d'intelligence est non seulement de mesurer les capacités de jugement et de recul cognitif (*critical thinking*), mais la capacité d'inhiber le système automatique et d'activer le système manuel. Cette incapacité d'inhibition explique pourquoi certaines personnes au demeurant fort intelligentes peuvent produire des jugements sur ce qu'il convient de faire ou de croire qui frappent par leur irrationalité.

Cela dit, la difficulté d'inhiber des processus automatiques de gestion des comportements est typiquement vue par les philosophes contemporains comme des cas de compulsion ou de dépendance, dans lesquels l'agent a certes un contrôle moteur orienté vers un but – et à ce titre, ils relèvent d'action –, mais pas de contrôle conscient[55]. C'est pour cette raison que ces comportements ne sont pas décrits comme des cas de stricte akrasie, ni même de

[55] Audi, par exemple, insiste sur la distinction entre la faiblesse *de* la volonté (akrasie) avec la faiblesse *dans* la volonté (compulsion dont l'agent veut consciemment se défaire) :

> « *A low level of strength in the overall faculty of will, such that, as compared with people of stronger will, there are more deeds (or more deeds of relevant kinds) which one can be compelled to do even if one judges one should not. […] Someone sufficiently strong-willed exhibits neither kinds of weakness ; but one can suffer from weakness in the will without being incontinent : circumstances being favorable, volitional weakness as a feature need not be reflected in volitionally weak action* ». (1993 : 322).

faiblesse de volonté au sens large du terme. Or, l'incapacité d'inhiber des processus décisionnels (ou cognitifs) ne relève pas toujours de compulsion. Si je ne m'engage pas dans la bonne sortie sur l'autoroute, par exemple, parce que je suis trop habitué de continuer mon chemin, je n'agis pas nécessaire de manière compulsive. Je n'ai simplement pas porté suffisamment attention à ce que je faisais – ce qui m'aurait permis d'inhiber mon système automatique de prise de décision.

5.1.2 Le processus contrôlé de prise de décision : la délibération

Il est assez trivial de dire que les processus décisionnels sont commandés par des problèmes décisionnels. Les agents enclenchent de tels processus pour résoudre un problème qui se pose à eux. Il n'est pas facile toutefois de définir techniquement ce qu'est un problème décisionnel. Nous avons des intuitions à leur endroit. Nous les imaginons comme une croisée de chemins, entre lesquelles on doit choisir pour atteindre sa destination. Mais nous n'avons pas de critères comme tels pour déterminer si l'on a affaire à un problème décisionnel ou non. Les problèmes sont vaguement définis en fonction de nos buts, des solutions disponibles, ainsi que du degré de leur évidence. Cela est cependant une base suffisante pour décrire la nature et la structure de processus décisionnels.

Comme les solutions, dans un problème décisionnel, ne sont pas toujours évidentes, l'agent doit à l'occasion enclencher un processus de computation consciente des informations pertinentes. Un processus contrôlé vise à pallier ce manque de caractère évident au moyen de toutes sortes d'inférences. Cela ne le distingue pas cependant des processus automatiques. Par exemple, Steven Pinker a fourni une analyse éclairante du processus par lequel l'esprit infère des modèles mentaux tridimensionnels à partir de prémisses bidimensionnelles fournies par notre appareil perceptuel (Pinker, 1997). Conduire une automobile requiert également une série de décisions automatiques complexes reposant sur des anticipations d'événements non réalisés, et pas seulement sur des réactions à l'égard des événements qui se produisent en temps réel. La manière dont les agents utilisent des processus contrôlés de prise de décision indique qu'ils vont pareillement au-delà des prémisses fournies par leur appareil perceptuel.

Le processus contrôlé de prise de décision par excellence est la *délibération*. Lorsqu'un agent délibère, même rapidement, il élabore et examine mentalement des scénarios

alternatifs qu'il n'observe pas à ce moment[56]. La délibération, comme la plupart des processus qui engendrent des représentations mentales, part de prémisses et en infère un certain nombre de conclusions.

Par exemple, depuis Aristote, on conçoit communément les solutions à des problèmes décisionnels comme l'usage de moyens appropriés pour atteindre une fin – qu'on peut également décrire comme des buts, des objectifs, des desseins, des résultats souhaitables, des intentions, des vues, des visées, des motifs, cibles, etc. On ne peut trouver de solution à un problème décisionnel que si l'on a une idée de la fin qu'on souhaite atteindre. Si l'on admet qu'une solution adéquate ou une solution *tout court* dépend à la fois de la variable *fin* et de la variable *moyen*, les agents doivent considérer que la variable *fin* est en partie fixée au départ – même si c'est de manière temporaire et relativement floue dans ses détails – sans quoi le processus décisionnel tournera à vide[57].

---

[56] Bien qu'il n'y ait pas de consensus sur le sujet, beaucoup de raisons militent en faveur de l'hypothèse que les mécanismes qui président à l'élaboration des scénarios alternatifs sont essentiellement ceux de l'imagination (Laird, 2009; Byrne, 2005). L'imagination, sorte de simulateur interne des choix, a un avantage adaptatif considérable. Il permet aux organismes qui en sont capables de tester des décisions sans risquer d'être éliminés. Karl Popper (1979) avait bien saisi ce trait lorsqu'il affirmait que le propre des hypothèses scientifiques est qu'on peut les laisser mourir à notre place.

Les scénarios n'ont pas besoin d'être identiques dans leurs détails aux scénarios qui se produiraient réellement si l'on optait pour le choix scénarisé. Seuls certains aspects pertinents ont une correspondance isomorphique. Les aspects isomorphiques peuvent variés en fonction de la nature du problème décisionnel abordé par l'agent. Les problèmes d'orientation dans un environnement physique qui comporte des portions qui sont à l'extérieur du champ visuel de l'agent requièrent l'usage d'une carte mentale si l'agent ne dispose d'aucune carte réelle. Or, Kosslyn (1980) a découvert que les cartes mentales possèdent des propriétés isomorphiques étonnantes. Lorsqu'on demande aux sujets d'une expérimentation de fixer leur attention de manière alternative sur divers points identifiés sur une carte réelle qu'ils ont préalablement examinée, on se rend compte que le temps nécessaire pour fixer leur attention est proportionnel à la distance réelle qui sépare ces points sur la carte.

Un scénario sera élaboré en fonction de ces aspects et d'un arrière-plan de connaissances factuelles et causales tacites (Dennett, 1995). Par exemple, voyant l'élément chauffant d'un poêle devenir rouge, je peux me demander si je dois y apposé la main – quelle drôle d'idée! J'imaginerai alors un scénario où je touche à l'élément et me brûle la main, parce que je connais des faits pertinents relatifs aux éléments rougis et à ce que peut faire la chaleur intense sur des tissus corporels. Or, la couleur du poêle et sa marque de commerce ne sont pas assez pertinentes pour les intégrés au scénario que je me fais. Ma simulation mentale s'alourdirait trop si je les intégrais par souci de réalisme.

D'ailleurs, le caractère économique du processus d'élaboration des scénarios imaginaires ne se manifeste pas qu'au niveau du degré de détails et de résolution. Ruth Byrne a décrit dans le détail ce qu'elle appelle les conditions cognitives de la création d'alternatives à la réalité. Le caractère limité de la mémoire de travail se reflète clairement dans le processus d'élaboration des scénarios, et cela, à plusieurs endroits.

Les agents ne peuvent garder à l'esprit qu'un nombre limité de scénarios hypothétiques et allègent leur mémoire en n'examinant que les plus probables ou en ne modifiant qu'un seul paramètre d'un environnement conçu, par ailleurs, comme étant tout à fait stable. Cette stabilité putative de tous les autres paramètres et les variables de l'environnement imaginé est, pour des raisons de limitation de la mémoire de travail, typiquement exagérée. C'est un facteur important d'inadéquation de certains scénarios selon Byrne (2009 : 15-40.).

[57] Cette manière de concevoir les problèmes décisionnels ne nous oblige pas à adhérer à une conception purement instrumentale de la rationalité pratique. On peut soutenir du coup qu'une solution adéquate à un problème décisionnel ne repose pas exclusivement sur le choix des moyens, mais également sur le choix des fins. Aussi, d'une manière générale solutionner un problème décisionnel pourrait revenir à utiliser un moyen adéquat pour atteindre une fin également adéquate. Tout dépend des contraintes qu'on souhaite donner à notre théorie du choix rationnel pour la catégorisation des « solutions ».

Certains ont critiqué le modèle du raisonnement pratique d'inspiration aristotélicienne parce qu'il n'intègre pas le concept de scénarios alternatifs. Pour Jaakko Hintikka, par exemple, cela a des répercussions importantes pour le débat sur l'existence de la stricte akrasie. En effet, la stricte akrasie implique que les agents élaborent mentalement au moins deux scénarios alternatifs dont l'un est conforme à la droite raison et l'autre akratique. Or, s'il n'y a pas au départ *simultanément* deux alternatives représentées mentalement, alors il ne peut y avoir de stricte akrasie (Hintikka, 1973).

Cela dit, rien ne s'oppose à l'idée d'intégrer des raisonnements moyen/fin à l'intérieur d'un schème délibératif portant sur des alternatives. C'est d'ailleurs ce que les agents font, ne serait-ce qu'implicitement, lorsqu'ils délibèrent.

5.1.2.1 Former un meilleur jugement ou une préférence synthétique complète : mode d'emploi

La tradition philosophique et la sagesse populaire voient dans le processus de délibération l'une des expressions abouties de notre nature rationnelle. La prudence veut que nous examinions l'ensemble des scénarios praticables avant de faire un choix qui devrait, au final, refléter une considération de tous les aspects pertinents de ceux-ci. Les décisions rationnelles ne se prennent pas de manière automatique, mais sont censées couronner un processus de délibération complète.

Le résultat d'une délibération est conceptualisé différemment selon qu'on se trouve dans la tradition philosophique cognitiviste ou dans la tradition hédoniste. Dans la tradition cognitiviste, une délibération réussie aboutie à la formation d'un meilleur jugement, tandis qu'elle correspond à la formation d'une préférence complète dans la tradition hédoniste. Bien qu'il s'agisse de deux schèmes conceptuels distincts, on y conçoit la fonction de la délibération de manière semblable : identifier le meilleur choix en maximisant certaines valeurs – de nature hédonique ou non. Mais comment y arrive-t-on?

Par définition, la délibération est un processus computationnel contrôlé qui consiste à peser le pour et le contre des alternatives qui s'offrent à nous. La tradition philosophique s'est souvent contentée d'affirmer que la délibération est une sorte de balance qui nous permet d'identifier les meilleurs choix étant donné les informations souvent partielles que nous avons, sans décrire dans le détail les étapes du processus. De manière analogue, on

conseille à l'occasion à un ami aux prises avec un problème décisionnel important de bien prendre le temps d'examiner les alternatives avant de faire un choix et de choisir ce qui pèse le plus lourd dans la balance. Mais on ne lui fournit, comme telle, aucune méthode explicite pour le faire. On prend pour acquis que cela se fait naturellement et de manière plus ou moins inconsciente – et d'ailleurs, on serait bien en mal d'expliciter ce processus qu'on estime pourtant à la base des décisions rationnelles.

La description des raisonnements pratiques proposée par Aristote ne fournit qu'une seule pièce du puzzle délibératif. En fait, tout ce que le raisonnement pratique montre se résume dans l'idée que l'on doit choisir en fonction d'un quelconque but ou d'une chose dont on souhaite éviter la réalisation. Aristote ne nous donne pas de clefs pour choisir entre des alternatives. Étrangement, une partie de la réponse provient historiquement d'une lettre que Benjamin Franklin adressa à son neveu, le chimiste Joseph Priestley[58]. Franklin explique comment on doit identifier le meilleur choix dans un problème décisionnel concret comme celui auquel Priestley tente de trouver une solution :

> *« In the Affair of so much Importance to you, wherein you ask my Advice, I cannot for want of sufficient Premises, advise you what to determine, but if you please I will tell you how.*
>
> *When these difficult Cases occur, they are difficult chiefly because while we have them under Consideration all the Reasons pro and con are not present to the Mind at the same time; but sometimes one Set present themselves, and at other times another, the first being out of Sight. Hence the various Purposes or Inclinations that alternately prevail, and the Uncertainty that perplexes us.*
>
> *To get over this, my Way is, to divide half a Sheet of Paper by a Line into two Columns, writing over the one Pro, and over the other Con. Then during three or four Days Consideration I put down under the different Heads short Hints of the different Motives that at different Times occur to me for or against the Measure. When I have thus got them all together in one View, I endeavor to estimate their respective Weights; and where I find two, one on each side, that seem equal, I strike them both out: If I find a Reason pro equal to some two Reasons con, I strike out the three. If I judge some two Reasons con equal to some three Reasons pro, I strike out the five; and thus proceeding I find at length where the Balance lies; and if after a Day or two of farther Consideration nothing new that is of Importance occurs on either side, I come to a Determination accordingly.*
>
> *And tho' the Weight of Reasons cannot be taken with the Precision of Algebraic Quantities, yet when each is thus considered separately and comparatively, and the whole lies before me, I think I can judge better, and am less likely to take a rash Step; and in fact I have found great Advantage from this kind of Equation, in what may be called Moral or Prudential Algebra.*

[58] Avant lui, Darwin (1887 : 232-233) avait abordé le problème en explicitant les raisons qui devraient, selon lui, lui permettre de trancher la question de savoir s'il devait ou non se marier. Toutefois, sa description du processus délibératif est moins détaillée et plus « expressionniste » que celle de Franklin.

Wishing sincerely that you may determine for the best, I am ever, my dear Friend.
Benjamin Franklin (1706-1790) »

L'*algèbre prudentielle* de Franklin est une méthode de délibération qui consiste donc à rechercher toutes les raisons, positives et négatives, pertinentes à la solution d'un problème décisionnel, les additionner de manière à voir de quel côté la balance penche. La méthode ne semble adaptée qu'aux dilemmes décisionnels. Or, si on affaire à une série comportant plus de deux alternatives, il est toujours possible de ramener le problème à un problème de choix binaire unique ou de diviser le problème initial en séquences de sous problèmes binaires. Par exemple, au lieu de me demander si je dois prendre mes vacances dans Charlevoix, à Paris ou à Bruxelles, je peux me demander si je dois prendre mes vacances au pays ou à l'étranger. Si la balance penche du côté de l'étranger, alors je comparerai les avantages/désavantages relatifs de Paris et Bruxelles – toujours en utilisant la méthode de Franklin.

De manière plus technique, la méthode de Franklin est une procédure qui consiste à soupeser de manière linéaire les combinaisons de raisons : peser et additionner. Il y a d'autres procédures linéaires apparentées à celle de Franklin.

Le psychologue Egon Brunswick a également proposé dans les années 1950 une variante de la méthode de Franklin qui était plus près du processus computationnel réel des agents délibérants. La méthode de régression linéaire plus sophistiquée consiste à tenir compte du poids probabiliste des raisons dans la délibération. Par exemple, si vous faites face à une personne dont vous devez déterminer l'âge, vous allez chercher un certain nombre d'indices pertinents : la couleur des cheveux, la prépondérance des rides, sa posture, son goût vestimentaire, etc. Or, si tous ces indices sont pertinents, certains sont de meilleurs indicateurs que d'autres. Idéalement, votre jugement devra être non seulement basé sur ces indices, mais vous devrez également peser, par le biais d'une analyse régressive, le poids probabiliste de chacun de ceux-ci en isolant les signaux qui sont de bons indicateurs (Hastie & Dawes, 2010). Par exemple, si je dois choisir entre Maryse et Denise, et que la fidélité est un critère important, alors j'isolerai des indices de fidélité et d'infidélité potentielles comme l'absence de rapport d'infidélité provenant des proches, l'intérêt marqué pour les autres hommes, etc.

Mais il y a également une procédure linéaire plus simple et plus rapide que l'algèbre de Franklin proposée Robyn Dawes à la fin des années 1970. Dawes soutient que les agents prennent la plupart du temps de bonnes décisions s'ils ne font que faire une liste mentale des raisons positives pour soustraire ensuite celle-ci l'ensemble des raisons négatives (1979 : 95-106). L'alternative qui se trouve avoir le plus de raisons positives ou le moins de raisons négatives est celle qui l'emporte. Cette méthode est connue sous le nom de la *règle de Dawes*.

Peu importe la méthode qu'on utilise cependant, les raisons qui permettent de soupeser les alternatives ne doivent pas être fortement corrélées. Par exemple, si je dois choisir entre deux automobiles et que, *entre autres*, leurs propriétés (relationnelles) esthétiques m'importent, je n'inclurai pas dans ma délibération l'ensemble des raisons suivantes : « est plus belle que », « est plus attrayante que », « est plus jolie que », « est plus gracieuse que » et « a un meilleur look que ». Toutes ces raisons se recoupent trop pour toutes les inclure dans ma délibération. Je devrai en choisir une seule, sans doute celle qui ressort du lot comme étant la plus générale ou qui recoupe au maximum les autres.

La délibération telle que Franklin la conçoit est compatible avec n'importe quelle conception fondamentale de la motivation. On peut accepter l'idée que Franklin a rendu explicite l'essentiel du processus délibératif et considérer que les raisons computables relèvent d'une estimation de l'importance réelle des caractéristiques propres à chacune des alternatives. De même qu'on peut l'accepter, mais considérer que les raisons computables relèvent plutôt d'une appréciation hédonique des conséquences et des moyens qu'impliquent chacune des alternatives. Dans le premier cas, on estimera que le processus consiste à computer des jugements partiels en vue de l'élaboration d'un meilleur jugement (tradition cognitiviste). Dans le second cas, on estimera que l'agent délibérant synthétise à partir des préférences partielles une préférence complète censée résumer le résultat de sa délibération.

5.1.2.2 Deux modèles de comparaisons dans la délibération

L'algèbre prudentielle de Franklin n'a d'algébrique que le nom. En fait, Franklin a explicité une procédure de computation des raisons et non des valeurs. À la suite de Sharif, Simonson et Dversky, on peut distinguer le modèle de computation *basée sur les raisons* du

modèle *basé sur des valeurs* en exploitant des ressources de calcul et de pondérations fines (Sharif, Simonson & Tversky, 1993). Dans un modèle basé sur les valeurs, la notion d'utilité est non seulement centrale, mais est l'étalon unique de toute comparaison. Les valeurs numériques représentent des quantités d'utilités espérées. Un modèle basé sur les valeurs permet de pondérer finement les conséquences (et même les moyens) qu'impliquent le choix d'une alternative et d'estimer précisément leur probabilité de réalisation.

Si, par exemple, je souhaite obtenir une paire de souliers, je devrai examiner différents scénarios d'achat. Je comparerai des souliers en tenant compte de plusieurs critères partiels tels que le niveau de confort, la durabilité, le caractère tendance ou ringard, et le prix. Pour chacun de ces critères partiels, je coterai chacune des paires examinées sur une échelle de 1 à 10. Pour chaque paire de souliers, j'opérerai une sommation de la valeur de chacune de ses caractéristiques et établirai sur cette base un ordonnancement complet des scénarios. Je serai de cette manière en mesure d'identifier celui qui est préférable entre tous. Évidemment, les critères partiels peuvent avoir des poids différents. Si je privilégie le confort sur toute autre considération, alors je devrai pondérer ce critère en conséquence. Par exemple, si j'estime que le confort est à peu près deux fois plus important pour moi que chacune des autres caractéristiques prises individuellement, alors je pourrai mesurer le confort à l'aide d'une échelle de 1 à 20, ou même de -20 à 20 si j'estime qu'un manque de confort représente un inconvénient majeur, ou même multiplier la valeur correspondant aux poids relatifs des scénarios et en faire ensuite la sommation.

Les probabilités de réalisation des conséquences sont un facteur de choix important. On peut difficilement identifier le meilleur choix si l'on n'a pas une certaine idée des risques et des probabilités de réussite. Or, non seulement leur estimation peut se faire de manière beaucoup plus précise dans un modèle basé sur des valeurs que dans le modèle concurrent, mais, chose plus importante, il est beaucoup plus facile de pondérer les poids respectifs des conséquences *et* de leur probabilité. En effet, depuis Bernoulli (1700-1782), beaucoup de théoriciens adhérant au modèle basé sur des valeurs considèrent que l'utilité espérée d'une alternative doit être le produit mathématique de l'intensité du désir pour l'ensemble de ses conséquences et du degré de certitude subjective de leur réalisation.

Cela dit, le modèle *basé-sur-des-valeurs* reste difficile à appliquer consciemment dans un monde complexe et échoue certainement à capter des aspects importants de la délibération

des agents dans des situations réelles (Sharif, Simonson & Tversky, 1993). Il est particulièrement mal adapté à des situations où les options alternatives sont également attirantes, difficiles à évaluer[59] ou incommensurables[60], où les probabilités de réalisations des conséquences sont difficiles à déterminer ou carrément inconnaissables[61]. C'est pourquoi on peut lui préférer le modèle *basé-sur-des-raisons*, plus souple – en fait, on peut considérer à peu près n'importe quoi comme une raison –, mais plus vague parce qu'essentiellement de nature qualitative.

5.1.3 Le résultat de la délibération et le début du problème akratique

Bien que j'aie donné une description très sommaire du processus de délibération, il correspond dans ses grandes lignes assez bien, je crois, à la manière dont les philosophes et théoriciens du choix rationnel l'envisagent. C'est essentiellement de cette manière que les agents, maximisateurs de nature, arrivent à identifier le meilleur choix – ne serait-ce que de manière plus ou moins inconsciente. Une fois que l'agent a formé un jugement à l'égard du meilleur scénario disponible pour lui, étant donné les informations dont il dispose, il n'a qu'à le réaliser.

En fait, c'est habituellement ici que les philosophes isolent un problème de faiblesse de volonté, et non en amont dans le processus de délibération. Conformément à la conception Standard, vivre un épisode akratique, c'est décider à l'encontre de ce qu'on estime pourtant

---

[59] Même quand on peut faire un bilan *a posteriori*, il est souvent impossible de quantifier les gains et les pertes véritables des choix. Par exemple, il est difficile de quantifier les gains et les pertes de comportements négligents comme ceux qui ont causé le déversement de millions de barils de pétrole brut dans le Golf du Mexique pendant l'été 2010. Il ne fait pas de doute que bien que les coûts furent astronomiques, mais chiffrables pour la pétrolière B.P., il est difficile d'évaluer le coût total pour l'environnement et pour les générations futures que représente cette catastrophe. Pour ce qui est des gains, cela a permis de conscientiser davantage les gens aux pratiques de l'industrie pétrolière et à la fragilité des écosystèmes. Mais jusqu'à quel point? Est-ce que cela sera suffisant pour prévenir une catastrophe encore plus grande? On ne peut pas répondre à ces interrogations.

[60] Le problème de l'incommensurabilité fut d'ailleurs considéré par Sidgwick (1874) comme l'un des principaux écueils pour toute théorie de la délibération rationnelle. Un exemple d'incommensurabilité est le dilemme auquel fait face la plupart des étudiants : Dois-je opter pour un appartement situé à 20 minutes du campus, mais est 80$ plus cher, ou pour celui à 40 minutes du campus?

[61] Les théories de la décision en économie distinguent le risque de l'incertitude. Le risque est une probabilité d'échec connu de l'agent alors que l'incertitude relève d'une variable inconnue de l'agent. Dans certaines situations, une recherche approfondie permet de transformer les incertitudes en risque. Mais dans beaucoup de situations, l'incertitude est plus robuste et l'agent n'a d'autre choix que de se contenter d'une vague approximation de l'étendue des risques que comportent les scénarios praticables. Or, même ces approximations ne constituent pas toujours des avenues que l'agent peut emprunter en dernier recours. Pour reprendre une expression utilisée par Donald Rumsfeld, secrétaire à la défense des États-Unis sous le mandat de George W. Bush. Le problème des décisions politiques à prendre dans un contexte international aussi compliqué que le nôtre est que, non seulement il y a des inconnues, mais également des *inconnues inconnues*! (2002).

être le mieux. Cela *suppose* que l'agent a dû faire face à un problème décisionnel, qu'il a su inhiber son système automatique de prise de décision et enclencher un processus de délibération, lequel a finalement abouti à la formation d'un jugement ou d'une préférence à l'égard du meilleur scénario réalisable par lui.

Décider à l'encontre de son meilleur jugement ou une préférence complète a quelque chose de paradoxal parce que l'agent qui entame et qui mène à bien un processus de délibération rationnel, le fait *parce qu*'il fait face à un problème décisionnel sans solution évidente et tente de trouver une solution en enclenchant ce processus qui améliorera au mieux sa condition ou la dégradera le moins. C'est comme si nous faisions face à un problème pratique que seul un outil nous permettrait de résoudre, que nous nous procurions l'outil, mais refusions de nous en servir. L'agent akratique est, selon la conception Standard, celui qui échoue à être suffisamment motivé par le résultat de sa délibération.

5.2 Conclusion

Il est nécessaire de prendre en compte le modèle du double processus sans quoi on surdiagnostiquerait des épisodes de stricte akrasie. La capacité des agents à inhiber leur processus automatique et à enclencher un processus contrôlé de délibération rationnelle est une condition pour la formation d'un meilleur jugement *et* son exécution. Bien que le processus ne mène par toujours à la formation d'un jugement *approprié*, l'agent tient minimalement compte de ses capacités, de ses croyances et de ses désirs (ou préférences) et émotions dans l'évaluation des alternatives qui s'offrent à lui. Avec ces considérations en arrière-plan, le problème que pose la stricte akrasie dans le débat philosophique devrait être celui de savoir si le résultat de la délibération motive *à tout coup* l'agent à prendre une décision s'y conformant *alors qu'il contrôle consciemment le processus menant à la prise de décision et l'exécution de celle-ci*.

J'aborderai dans le prochain chapitre porte sur le débat classique qui oppose les partisans de l'internalisme à ceux de l'externalisme à l'égard du lien entre le meilleur jugement et la décision. Après avoir exposé les positions opposées et une position mitoyenne – que je vais, du reste, critiquer dans le détail –, je vais exposer des raisons empiriques solides et décisives de croire que le lien entre le meilleur jugement et la décision est externe et non interne. Même si on adhère au modèle du double processus et qu'on exclu les cas où les

agents ne sont pas en mesure d'inhiber leurs systèmes automatiques d'exécution de décision, il resterait beaucoup trop de contre-exemples à la conception internaliste pour que celle-ci soit viable.

6
L'internalisme et l'externalisme

In any moment of decision the best thing you can do is the right thing, the next best thing is the wrong thing, and the worst thing you can do is nothing.

–Théodore Roosevelt

6.1 La connexion entre le meilleur jugement et la décision

Le débat philosophique contemporain au sujet de l'existence de la stricte akrasie tourne autour de la nature du lien entre le meilleur jugement de l'agent et la décision de s'y conformer. Bien que le débat porte sur un problème de motivation comme tel, il a surtout eu lieu – et continue à ce jour de l'être – dans le domaine de la théorie morale (William, 1979, 1980 ; Korsgaard, 1986 ; Railton, 1986 ; Smith, 1994). Les théories morales ne sont pas toujours éclairantes à ce propos parce qu'elles ont tendance à traiter les considérations normatives d'ordre moral comme ayant un statut particulier au sein de l'espace motivationnel de l'agent. En fait, les discussions portent plus sur l'internalisme *normatif* que *motivationnel*. L'internalisme motivationnel concerne la nature de l'explication des décisions réelles des agents, alors que l'internalisme normatif concerne la force motivationnelle des raisons que l'agent *devrait* avoir. Le débat entre internalistes et externaliste motivationnel porte sur la nature de la connexion entre *ses* raisons réelles et sa décision. Par contraste, dans le cadre du débat entre internalistes et externalistes normatifs, on se demandera, par exemple, si l'on peut condamner les choix d'Hitler alors qu'il n'était pas du tout sensible au sort des peuples dont il a exigé la persécution. Cela ne concerne pas directement le problème motivationnel que représente la faiblesse de volonté. Qui plus est, ces théories proposées dans le cadre du débat entre internalistes et externalistes normatifs identifient souvent la faiblesse de volonté comme une sorte de cas limite, et estiment qu'une force de volonté minimale constitue le pré requis à l'établissement d'une connexion entre le jugement moral de l'agent et la décision de s'y conformer.

Les théoriciens qui estiment que le meilleur jugement motive *nécessairement* et/ou de *manière suffisante* les agents à prendre une décision s'y conformant sont réputés adopter

une conception internaliste de la motivation. Les théoriciens défendant plutôt l'hypothèse que ce lien est *contingent* et peut être à l'occasion *insuffisant* adoptent une conception externaliste. Il y a des variations importantes entre les théories et les manières d'envisager le problème, mais elles portent toutes essentiellement sur la nature du lien entre le résultat de la délibération et la décision de l'agent.

Dans le cadre de la conception internaliste, on peut accepter et on *doit* accepter l'idée que les agents puissent décider à l'encontre de leur meilleur jugement lorsqu'ils pilotent leurs activités à l'aide de leur système automatique. On peut également accepter l'idée qu'il puisse y avoir un délais important entre le moment où un meilleur jugement est formé et le moment où l'agent prend effectivement une décision motivée par celui-ci. Mais si l'agent ne dispose pas de meilleures raisons pour décider autrement, il se conformera à son meilleur jugement. Il n'y a donc pas, pour un internaliste, de stricte akrasie – seulement des changements d'avis plus ou moins rationnels. En revanche, elle demeure toujours une possibilité pour un partisan de la thèse de l'externalisme motivationnel.

J'exposerai tour à tour les arguments du philosophe Richard Hare, qui a exprimé et défendu un internalisme motivationnel fort, et ceux de l'externalisme d'Alfred Mele. Je présenterai et critiquerai plus en détail l'internalisme faible de Donald Davidson parce qu'il représente sans doute *la* position de référence dans tout ce débat, mais j'exposerai ensuite quelques éléments critiques généraux sur les termes mêmes du débat.

6.1.1 L'internalisme fort de Hare

Bien qu'il ne soit pas à la source du label – dû au philosophe W. D. Falk (1948) – Richard Hare a formulé le premier une position élaborée et un argument clair dans les années 1950 et 1960 en faveur de l'hypothèse internaliste. Pour Hare, former le jugement *que je dois faire* x revient à formuler à mon endroit un *impératif* qui a une force motivationnelle suffisante pour *que je décide de faire* x.

Le scepticisme de Hare à l'égard de la stricte akrasie repose sur l'idée que les jugements évaluatifs – comme le sont les préférences – entretiennent une connexion particulière avec l'action *que n'ont pas les jugements factuels*. Les premiers ont le caractère spécial d'une directive, aussi partielle soit-elle, pour la conduite (Hare : 1952 ; 1963). On ne peut pas répondre de manière satisfaisante à la question : « Que dois-je faire? », en ne relatant que

des faits. On doit offrir dans sa réponse un commandement ou un impératif de nature prudentielle (ou morale). Or, nous dit Hare, c'est justement la fonction des commandements et des impératifs que de servir de guide pour l'action. Si un agent s'engage dans une ligne de conduite particulière, alors il y a un jugement évaluatif qui lui commande ou lui ordonne de le faire. Si cet agent est sincère lorsqu'il rapporte son jugement, alors ce jugement concordera à ce qu'il essaie effectivement de faire, parce que c'est une sorte de : « […] *tautologie que de dire que nous ne pouvons pas asserter* […] *un commandement que nous nous adressons sans en même temps l'exécuter, si on a l'occasion de le faire et s'il est en notre pouvoir de le faire.* » (1952 : 20)

Hare et ses commentateurs ont d'ailleurs fourni une version sémantique de cette position, soit pour la critiquer ou pour la développer. La version sémantique de l'internalisme motivationnel est l'*impérativisme*. L'impérativisme stipule qu'un énoncé tel que « *je dois faire* x » entraine logiquement l'impératif « *fais* x »[62]. Mais je m'en tiens ici à la version substantielle ou empirique de la thèse de Hare et non à sa version sémantique.

Cela dit, Hare fournit tout de même une explication des cas putatifs de stricte akrasie. Pour lui, les agents que la tradition philosophique voit comme étant strictement akratique

(a) ne sont pas vraiment libres de décider de faire *x* parce qu'il y a trop de contraintes, ou

(b) ne sont pas vraiment sincères, ou sont carrément hypocrites, lorsqu'ils affirment qu'il est préférable de faire *x* alors qu'ils décident de faire *y* (comme dans la mauvaise foi sartrienne), ou

[62] Ce genre d'analyse sémantique a cependant une portée très limitée. Premièrement, l'analyse des énoncés ne nous apprend pas grand-chose sur les phénomènes ou événements décrits par ceux-ci. Deuxièmement, une analyse sémantique présuppose ou implique qu'on a déjà réglé certaines questions factuelles concernant notamment l'existence de certaines choses, alors qu'elles sont l'objet du débat. Pour une discussion de l'impérativisme, voir Peter Geach (1965) et Tappolet (2004 : 5-12).

Cela dit, Joseph Heath a formulé un argument sémantique en s'appuyant sur certaines remarques formulées par Sellars (1963). Dans sa critique de la position externaliste, Heath cite le dicton de Sellars suivant lequel, c'est « […] *une vérité nécessaire que les personnes tendent de faire ce qu'elles jugent devoir faire, et c'est aussi une vérité nécessaire que les personnes qui occupent une position linguistique de laquelle elles signifient* je dois faire A*, tendent à faire A.* » (2008 : 229)

Heath développe assez peu cet argument – souhaitant mettre plutôt l'emphase sur l'existence d'explications alternatives de la faiblesse de volonté comme l'instabilité dynamique des préférences. Mais il estime qu'une expression comme « vouloir faire *x* » implique « essayer de faire *x* » dans nos usages linguistiques socialement enrégimentés. L'externalisme impliquerait, selon Heath, la possibilité d'un désalignement *systématique* entre ce qu'un agent affirme sincèrement devoir faire et ce qu'il fait effectivement. Or, la perspective d'un tel désalignement nous amènerait en fait à réviser nos attributions d'intention à cet agent, et cela, conformément aux règles langagières qui encadrent l'usage de ces expressions.

(c) ne mentionnent qu'une convention sociale ou une opinion de « *bon ton* », ou

(d) ont seulement le sentiment qu'ils devraient faire *x* et rapportent en fait verbalement ce sentiment, mais sans plus, ou

(e) vivent une sorte de division de leur moi dont une partie les pousse à s'engager dans une ligne de conduite qu'ils jugent mauvaise, ou

(f) ne réalisent pas vraiment la portée de leur affirmation – notamment concernant l'application de cette affirmation à leur propre cas (comme chez Aristote).

Une analyse fine des cas putatifs de stricte akrasie – comme celles qu'on retrouve dans les œuvres littéraires de Balzac et de Proust – devrait révéler, si on en croit Hare, qu'ils relèvent individuellement de l'une ou l'autre de ces possibilités. Cela ménage en partie nos intuitions que nous interprétons spontanément comme des intuitions portant sur des cas réels de stricte akrasie.

C'est un avantage important de la conception de la motivation de Hare. Toutefois, isoler et décrire ces possibilités ne milite pas vraiment en faveur de l'hypothèse internaliste. Leurs existences restent compatibles avec une position externaliste. En fait, l'argument principal qu'élabore Hare en faveur de l'internalisme est une sorte de preuve par l'absurde. Hare avance que nous ne pouvons pas souscrire à l'externalisme parce qu'il implique des choses que nous ne pouvons pas accepter parce que cela va à l'encontre de nos intuitions les plus fortes à l'égard de la motivation.

Selon Hare, l'externalisme est une position erronée, parce que

> « […] *selon le point de vue considéré* [l'externalisme]*, il n'y a rien de plus bizarre dans le fait de penser que quelque chose est la meilleure chose à faire dans les circonstances, mais ne pas la faire, que dans le fait de penser qu'il y a une pierre ronde dans les environs et en prendre une autre à la place… Il n'y a rien qui requiert une explication si je choisis ce que je considère comme étant la pire chose en laissant de côté ce que je pense être la meilleure.* » (1964 : 68-69)

Pour Hare, il est clair que les jugements évaluatifs ont une fonction motivationnelle que n'ont pas les jugements de fait. Même si l'on doit avoir des informations factuelles pour s'orienter dans une direction particulière, celles-ci ont toujours pour trame de font des évaluations portant sur des états du monde possible. Si, par exemple, j'essaie de convaincre un ami de ne pas acheter tel modèle de voiture parce qu'il n'est pas fiable, je juge

implicitement que mon ami préfère vivre dans un monde où, *ceteris paribus*, il possède une voiture fiable que dans un monde où il possède une voiture non fiable. Hare ne stipule pas, par contre, que la distinction entre fait et valeur (au sens large du terme) est une dichotomie. Les jugements de fait et de valeur sont situés aux pôles opposés d'un continuum. Des termes comme « bon », « valide », « fiable » ou « adéquat », par exemple, en viennent à avoir un contenu factuel ou appréciatif plus proéminent suivant les situations (1963 : 164).

Mais l'internalisme de Hare va un peu plus loin qu'impliquer que les jugements évaluatifs sont des ingrédients *nécessaires* pour motiver les agents à prendre des décisions. Il soutient que les jugements évaluatifs qui résultent d'une délibération ou qui sont situés au point de jonction avec la décision motivent *suffisamment* les agents à prendre une décision allant dans leur sens. C'est pourquoi on peut qualifie la position de Hare d'internalisme fort.

6.1.2 L'externalisme de Mele

Certains philosophes considèrent qu'il n'y a pas de lois psychologiques qui lient causalement et invariablement le meilleur jugement de l'agent et sa décision. Pour Alfred Mele, par exemple, il n'y aurait pas de connexions « nécessaires » entre le meilleur jugement de l'agent et sa motivation (suffisante) à prendre une décision particulière, même si l'agent effectue une délibération pleine et entière. Cette hypothèse relève de l'externalisme motivationnel et Mele est sans doute son plus féroce partisan. Pour Mele, nos intuitions les plus fortes militent en faveur de l'idée que le caractère évaluatif de nos désirs et leur force motivationnelle ne sont pas les deux faces d'une même médaille ou le reflet l'une de l'autre, mais que leurs poids respectifs peuvent croître *différemment* et même *inversement*. Dans une formule plus concise, Mele soutient la thèse que « […] *the balance of an agent's motivation can be* out of line *with it decisive better judgment event about something to be done here and now.* » (1987 : 84)

L'externalisme qui sert de repoussoir à Hare est une forme extrême que peu de théoriciens seraient prêts à défendre. On peut accepter l'hypothèse que des évaluations induites par des préférences, des désirs, des goûts, des souhaits, des espérances, etc. sont des constituants nécessaires pour motiver les agents à prendre des décisions, mais estimer qu'un meilleur jugement peut à l'occasion ne pas motiver suffisamment la personne qui l'a pourtant formé.

L'agent strictement akratique serait motivé par des considérations désidératives, et non seulement par des considérations cognitives (jugements de fait, estimation des probabilités, anticipation de la réalisation d'état hédonique futur, etc.). Aussi, il n'est pas requis d'adopter une forme d'externalisme extrême qui veut que les considérations désidératives ne soient pas nécessaires pour motiver une décision.

Mais pour Mele, les considérations désidératives sont toutefois traitées à deux niveaux différents dans l'esprit. Au premier niveau, elles servent d'« input » à la délibération et participent à la détermination de l'importance des raisons (leur poids) aussi bien que des alternatives et de l'ordre de priorité des objectifs à poursuivre. Au second niveau, elles fournissent l'impulsion nécessaire pour la production d'une décision. Au premier niveau donc, les considérations désidératives participent à l'évaluation des alternatives, alors qu'au second niveau elles motivent les agents à les choisir.

Le découplement du processus d'évaluation et de motivation explique pourquoi on peut vivre des épisodes de stricte akrasie. L'agent akratique a un problème de désalignement du processus d'évaluation du premier niveau avec celui du second niveau responsable de la motivation. Aussi, lorsque nous affirmons qu'un désir a un poids, cela peut signifier qu'il a une certaine priorité dans l'évaluation des scénarios réalisables pour un agent, mais cela peut également signifier qu'il a une certaine force motivationnelle susceptible de pousser un agent à prendre une décision particulière. Cette ambiguïté rendrait le phénomène de la stricte akrasie énigmatique. Mais si l'on accepte la proposition de Mele, alors le paradoxe se dissout.

Il reste par contre à expliquer pourquoi les agents éprouvent à l'occasion des difficultés d'alignement de leur processus d'évaluations. Pour expliquer le phénomène putatif de la stricte akrasie, Mele utilise certains résultats de recherches en psychologie de la motivation, et notamment la manière dont le psychologue George Ainslie (Ainslie, 2001 ; Ainslie & Haslam, 1992) les a interprétés, et quelques suggestions de la philosophe Amélie Rorty (1980).

Ainslie aurait, selon Mele, conceptuellement et empiriquement cerné un phénomène susceptible d'expliquer un nombre considérable de cas de stricte akrasie. Le *rapprochement temporel de consommations de biens ou de maux* modifierait considérablement notre

motivation à les obtenir ou à éviter d'obtenir en dépit du fait que ces consommations *n'exemplifient pas de nouvelles propriétés*. Par exemple, je peux avoir le lundi un mal de dent terrible que j'estime avoir pour cause une carie, et prendre par conséquent un rendez-vous pour une obturation prévue le vendredi. Je sais que l'obturation sera très déplaisante, mais j'ai si mal que je juge préférable de la subir. Or, j'annule mon rendez-vous le vendredi matin, prétextant un empêchement. L'explication de ce qui s'est passé est qu'*en dépit du fait que j'ai toujours autant de raisons de subir cette intervention*, j'ai perdu ma motivation à aller à mon rendez-vous parce que le point de consommation est beaucoup plus près de moi dans le temps qu'au moment où j'ai décidé de prendre mon rendez-vous.

Ce facteur temporel n'amène pas tous les agents (et dans toutes les situations) à prendre des décisions akratiques. Les stratégies de contrôle de soi comme l'engagement préalable (sorte de contrainte à soi-même analogue à celle qu'Ulysse utilisa pour éviter de tomber dans le piège des Sirènes), les paris privés et publics, les stratégies d'orientation de l'attention, etc. permettent de contrebalancer les effets de proximité temporelle en « nourrissant » ou « affamant » la motivation à s'engager dans certaines lignes de conduite.

Les suggestions de Rorty consistent en stratégies explicatives, en quelque sorte, complémentaires. Rorty identifie les *habitudes* et les *impetus* sociaux comme des causes possibles de l'akrasie. Mele soutient que l'implantation d'une mauvaise habitude modifie la motivation d'un agent à s'engager dans une ligne de conduite akratique, en dépit du fait qu'il a de bonnes raisons, à ses yeux, de ne pas s'y engager. Les *impetus* sociaux sont un type de facteurs externes qui entretiennent ou découragent des habitudes en fonction de son degré de conformité avec une pratique socialement partagée.

Mais la cause la plus importante pour Mele (également identifiée par Rorty) est une modification dans l'*attention* que les agents portent sur des points de consommation temporellement distants. Le fait de fixer son attention sur certains aspects d'un scénario de consommation et non sur d'autres a un effet considérable sur la motivation des agents à prendre la décision de les réaliser ou non. Mais fixer son attention sur un aspect ne revient pas à lui attribuer un poids (dans le sens évaluatif du terme) plus grand dans la délibération.

D'ailleurs, Mele soutient que les mécanismes de l'attention sous-tendent l'effet de proximité temporelle, et aussi les habitudes et les *impetus* sociaux. La proximité temporelle

produit une modification de notre motivation parce que plus un point de consommation est près de nous dans le temps, plus certains de ses aspects attireront notre attention. Pour ce qui est des habitudes, elles seraient conditionnées par des patterns rigides d'attention. Tandis que pour les *impetus* sociaux, on pourrait expliquer leurs effets sur la motivation par le fait qu'ils impliquent des récompenses et punitions socialement induites (implicitement autant qu'explicitement) plus ou moins rapprochées dans le temps.

La possibilité du désalignement de l'évaluation au niveau de la délibération et de celle qui produit effectivement la motivation à prendre une certaine décision requiert une conception externaliste du lien entre le meilleur jugement de l'agent et sa décision. Ce lien ne peut être suffisant, ni même nécessaire, parce qu'il dépend de facteurs contingents relevant de la direction que prennent les désirs motivants situés au second niveau.

6.1.3 L'internalisme faible de Davidson

La position de Donald Davidson est sans doute celle qui a suscité le plus de commentaires dans la littérature sur la faiblesse de volonté. Dans son article célèbre, *Comment la faiblesse de volonté est-elle possible?*, Davidson explique pourquoi on ne doit pas considérer, contrairement à la tradition philosophique, que les agents akratiques sont *logiquement* incohérents, même s'ils sont clairement *irrationnels*. En développant sa conception du problème de la faiblesse de volonté, Davidson tente de montrer qu'on peut agir à l'encontre de son meilleur jugement sans commettre de faute logique. Son argument est très sophistiqué, mais lui permet d'aménager une position théorique qui reflète un compromis entre l'internalisme et l'externalisme. À ce titre, il vaut la peine qu'on s'y attarde.

En premier lieu, Davidson soutient qu'il est indubitable que les comportements akratiques existent. Mais ce qui a pu faire croire à leur inexistence est qu'ils semblent incompatibles avec deux lois psychologiques tout aussi indubitables :

> L^1 : Si un agent veut faire *a* plus qu'il ne veut faire *b* et s'il se croit libre de faire *a* ou *b*, alors il fera intentionnellement *a* s'il fait soit *a* soit *b* intentionnellement.
>
> L^2 : Si un agent juge qu'il serait meilleur de faire *a* que de faire *b*, alors il veut faire *a* plus qu'il ne veut faire *b*.

Davidson ne nie pas la justesse de ces deux lois. Le fait qu'il souscrive à L² le range du côté des internalistes puisque cette loi établit une connexion entre le meilleur jugement évaluatif d'un agent et sa motivation à agir ou décider d'agir. Mais comment un agent peut-il alors aller à l'encontre de son meilleur jugement? La solution de Davidson pour concilier ces deux lois avec la faiblesse de volonté est de considérer que juger que *a* est meilleur que *b* n'équivaut pas logiquement à juger que *a* est meilleur que *b*, *toutes choses considérées*. Aussi, lorsqu'un agent agit ou décide à l'encontre de son meilleur jugement, il peut se conformer néanmoins à un jugement qu'une alternalive est préférable à une autre *tout court*.

À la première lecture, on a le sentiment que l'auteur essaie de nous mystifier. Mais en y regardant de plus près, on se rend compte que la distinction n'est pas spécieuse. Davidson développe son argument en identifiant les propriétés logiques de chacun de ces deux types de jugement préférentiel. C'est intéressant à plusieurs titres. Notamment parce qu'il tente de mettre, à mon avis, en relief certains aspects importants de la structure logique de la délibération rationnelle, et des procédures issues du modèle *basé-sur-les-raisons*, comme la méthode de Franklin ou la règle de Dawes.

Au cœur de la distinction de Davidson, on trouve la distinction entre les *jugements conditionnels* et les *jugements inconditionnels*. Il est donc nécessaire de comprendre cette dernière pour expliquer la possibilité de la faiblesse de volonté. Davidson met en relief le fait que les jugements conditionnels ont la forme logique « x *est préférable à* y *pour la raison* r´ ». Ces derniers jugements établissent un ordonnancement d'actions ou de scénarios relativement à une raison ou un critère particulier, ce qui n'est pas le cas du second type de jugement. La formation de jugements conditionnels ponctuerait l'ensemble de la délibération d'un agent. Celui-ci formera typiquement des jugements conditionnels *prima facie* relativement à une raison ou critère, puis formerait ensuite des jugements *prima facie* relativement à plusieurs raisons ou à plusieurs critères, jusqu'à l'obtention d'un meilleur jugement. Or, nous dit Davidson, le meilleur jugement est lui aussi un jugement conditionnel, puisqu'il a la forme logique : « x *est préférable à* y *pour la raison* (r_1, r_2,, ..., *rn*) », où (r_1, r_2,, ..., *rn*) représentent l'ensemble des raisons pertinentes pour l'agent.

Les jugements inconditionnels ont la forme logique « x *est préférable à* y ». Ils consistent comme les premiers en un ordonnancement d'actions ou de scénarios, mais pas

relativement à un ensemble de raisons ou de critères. Ils sont situés dans la délibération, à la toute fin, c'est-à-dire *au point de décision*. Dans toutes les situations où un agent prend une décision ou accomplit un acte intentionnel, il forme un jugement de ce type.

Si cette analyse est correcte, que peut-on dire d'un agent qui vit un épisode de faiblesse de volonté? L'akratès échouerait à se conformer à son meilleur jugement parce qu'à ce jugement ne correspond pas un jugement inconditionnel *au point de décision*. Un agent peut préférer *a* à *b* parce que *a* est sous un aspect r_1 plus avantageux que *b*, et préférer *b* parce qu'il est plus avantageux sous l'aspect r_2. Mais en considérant tous ces aspects – imaginons, par exemple, qu'il aurait utilisé la méthode de Franklin – il pourrait préférer *a*, mais choisir tout de même *b*. Dans ce cas, l'agent a échoué à établir que *a* est préférable à *b* inconditionnellement, ce qui l'aurait amené à se conformer à son meilleur jugement.

La distinction de Davidson lui permet de rejeter l'idée que l'akratès commet une faute logique en ne se conformant pas à son meilleur jugement. Comme on ne peut pas logiquement dériver un jugement inconditionnel d'un jugement conditionnel – même si celui-ci mentionne une infinité de conditions – on ne peut pas non plus dire qu'ils peuvent être logiquement incompatibles, ni dans les cas d'akrasie ni dans les cas de comportements rationnels. Or, même s'il n'y a pas de règle déductive qui nous permette de détacher un jugement préférentiel *tout court* d'un jugement préférentiel *toutes-choses-considérées*, il y a pour Davidson un principe rationnel qui enjoint de le faire. Pour éviter de tomber dans des schémas de décision akratique, nous devons appliquer systématiquement l'injonction « […] *d'accomplir l'action jugée la meilleure sur la base de toutes les raisons pertinentes disponibles.* » (1970 : 64)

C'est ce que Davidson appelle le principe de continence. Si on n'applique pas ce principe, on se comportera *de facto* de manière akratique. C'est ce qui fait que la position de Davidson relève d'une forme d'internalisme faible. Mais Davidson se rapproche de l'internalisme fort lorsqu'il soutient qu'il ne peut y avoir d'action intentionnelle accomplie à l'encontre d'un jugement inconditionnel qui nous oriente dans la *même* direction qu'un meilleur jugement.

Comment un agent peut-il juger que *a* est préférable, toutes choses considérées, à *b*, mais préférer toute de même *b* à *a*, tout court? Davidson rejette l'idée que l'agent akratique serait

victime d'une confusion logique et avance l'hypothèse que les causes du comportement akratique ne sont tout simplement pas des raisons. L'explication que donne Davidson de la stricte akrasie ne se résume pas à dire que les agents échouent à appliquer le principe de continence. Il fait un parallèle avec les croyances motivées pour expliquer sa possibilité :

> « *Si* r *est la raison que quelqu'un a de soutenir que* P*, alors le fait qu'il soutient que* r *doit être, je pense, une cause du fait qu'il soutient que* P*, mais, et c'est ici le point crucial, le fait qu'il soutient que* r *peut être la cause du fait qu'il soutient que* P *sans que ceci soit sa raison pour cela : et bien entendu, l'agent peut même penser que* r *soit une raison de rejeter* P. *» (1970 : 63)*

Une croyance motivée est une croyance à laquelle un agent adhère parce qu'il veut y adhérer ou parce que cela lui permettrait de solutionner un problème de dissonance cognitive. Nous estimons tous que nous ne devons pas croire quelque chose simplement parce que cela fait notre affaire[63]. Pourtant, les cas de croyances motivées sont réels. Il y a donc des états mentaux qui ne peuvent être des raisons pour adhérer à une croyance, mais qui néanmoins causent une adhésion à cette croyance. De plus, il est manifestement irrationnel d'entretenir des croyances motivées – ce qui rend le parallèle d'autant plus pertinent.

Davidson mentionne un exemple intersubjectif de causes d'origine mentale qui ne peuvent être des raisons. Je peux vouloir qu'une femme que j'affectionne, mais qui est indifférente à mon égard, se rende dans un restaurant où je compte aller. Aussi, je peux faire en sorte qu'elle croit que la personne qu'elle chérit le plus se trouvera dans ce restaurant. Or, si elle s'y rend, il est clair que mon désir a causé son action, mais qu'il ne constituait certainement pas pour elle une raison d'y aller.

---

[63] C'est d'ailleurs ce qui explique notre réaction spontanée au fameux pari de Pascal. Le pari de Pascal consiste à miser sur l'existence de Dieu parce que si on la refuse, alors on jouit certes davantage de moments discrétionnaires qu'on ne « gaspille » pas dans les lieux de culte, mais on risque la damnation éternelle. Tandis que si l'on mise sur Dieu, alors on perd la jouissance de ces moments, mais on ne risque pas la damnation éternelle. Un simple calcul montre qu'une probabilité plus grande que 0, aussi minime soit-elle, d'une souffrance éternelle, aussi légère soit-elle, pèse toujours plus lourd dans la balance que la probabilité de 1 de plaisirs temporellement limités, aussi intenses soient-ils. Lorsqu'on expose l'argument à des croyants ou à des incroyants, on assiste souvent à une réaction d'incrédulité.

L'argument ne convainc personne pour une raison particulière : il ne constitue pas une raison pour *croire en Dieu*, mais seulement une raison pour *vouloir y croire*. Or, on ne doit pas croire que *P* soit le cas, parce qu'on veut ou qu'on désire que *P* soit le cas, *tout court*. En fait, cela fut clair dans l'esprit de Pascal, et c'est pour cela qu'il proposait une méthode pour arriver à croire effectivement que Dieu existe, étant donné qu'on a une bonne raison, selon lui, de vouloir y croire. Calqué sur la méthode aristotélicienne pour devenir vertueux, Pascal soutenait que le meilleur moyen de devenir croyant est de se comporter comme un vrai croyant.

Cela permet à Davidson d'expliquer la possibilité de la faiblesse de volonté. Lorsque l'akratès prend une décision à l'encontre de son meilleur jugement, ce dernier constitue certes une raison pour décider autrement, mais la cause de son comportement réside dans le fait qu'il désire autre chose. L'akratès peut reconnaître que certains désirs ne constituent pas des raisons pour décider de s'engager dans une ligne de conduite particulière, comme il peut reconnaître que certains états mentaux ne sont pas des raisons pour croire en quoi que ce soit. Cette avenue théorique est jugée adéquate par Davidson parce qu'elle nous permet de relever un défi auquel doit faire face le théoricien qui veut expliquer le phénomène de la faiblesse de volonté, à savoir qu'il ne doit pas en mettant en relief les raisons de l'agent faire apparaitre celui-ci finalement rationnel (1991 : 115 ; 1985 : 347). L'akratès est irrationnel parce qu'il n'a pas vraiment de raison de faire ce qu'il fait, mais ses actions demeurent tout de même dirigées par des causes mentales[64].

6.2 Que penser du débat internalisme/externalisme?

Beaucoup de théoriciens en philosophie pratique ont pris part au débat en prenant position pour l'une des deux conceptions et en critiquant l'autre. Mais on peut émettre certains doutes sur les termes mêmes du débat ou sur la capacité des conceptions opposées de rendre compte des cas putatifs de stricte akrasie. Je mentionnai ici deux critiques qui, je le pense, sont assez sérieuses pour que l'on considère que ce débat ne doit pas être au cœur du problème de la faiblesse de volonté, et une critique de la notion de jugement inconditionnel qu'a proposée Davidson pour solutionner le problème.

6.2.1 Le statut causal de la connexion: une question de degré

Soutenir que les meilleurs jugements peuvent avoir une connexion interne ou externe avec les décisions (ou les actions) repose implicitement sur une conception causale particulière et simpliste de la motivation. Un partisan de l'internalisme soutiendra que le meilleur jugement entraîne causalement la réalisation d'une décision s'il est sincère et que l'agent est effectivement capable de le faire. Un externaliste soutiendra que cet effet n'est pas

[64] Cela permet également à Davidson de respecter son critère causal de l'agir, et d'éviter du coup de faire des comportements akratiques des actions non intentionnelles, compulsives ou pilotées exclusivement par le système automatique.

causalement nécessaire et requiert la réalisation de certains autres processus mentaux pour avoir réellement ce pouvoir causal.

Or, cela repose sur une confusion. On suppose, de part et d'autre, que, si l'on arrivait à prouver qu'une décision ou qu'une action réelle n'est pas conforme à un meilleur jugement sincère, alors cela impliquerait que celui-ci est causalement inerte et que, pour cette raison, on devrait le considérer comme étant externe, alors qu'en fait les jugements (peu importe où ils se situent dans la délibération) sont des *facteurs* de décision. Si un agent prend une décision ou agit à l'encontre de son meilleur jugement, cela n'implique pas que ce jugement n'affecte d'aucune façon la disposition de l'agent à prendre cette décision ou à agir à l'encontre de ce même jugement. Un agent qui forme un meilleur jugement augmente la probabilité qu'il prenne une certaine décision et diminue la probabilité qu'il en prenne une autre, *et cela reste vrai même s'il décide de ne pas s'y conformer*. Cela est analogue au fait que le tabac est un facteur de cancer même chez les fumeurs qui ne développeront jamais de métastases.

En tant que facteur décisionnel interne, le meilleur jugement a une connexion causale avec la décision, mais une connexion de nature probabiliste. S'il l'on tient à la distinction entre connexions internes et externes, alors on devrait l'envisager davantage comme une différence de degré dans la force des facteurs décisionnels relevant de la délibération.

D'ailleurs, même si un agent décide ou agit à l'encontre de son meilleur jugement, ce dernier continue d'avoir un impact réel et observable, et pas seulement sous forme probabiliste. Un agent strictement akratique prendra plus de temps pour prendre sa décision parce qu'il tentera de résoudre un conflit interne. Il prendra également plus de temps pour mettre à exécution sa décision, et agira de manière réticente ou se montrera hésitant. De plus, il sera davantage disposé à revenir sur sa décision que s'il n'a pas *ceteris paribus* formé un meilleur jugement allant dans la direction contraire.

Tout ceci nous indique que la connexion entre le meilleur jugement et la décision exemplifie des propriétés subtiles qui sont difficilement explicables dans le cadre du débat internaliste/externaliste.

6.2.2 L'akrasie comme problème situé en amont du meilleur jugement ou en aval de la décision

La faiblesse de volonté est traditionnellement vue comme une forme d'échec pratique qui ne résulte pas d'interférences externes, d'un manque de ressource ou d'un problème de coordination avec d'autres agents ou objet en mouvement, car c'est un échec dont la cause réside à l'intérieur de l'agent lui-même. Mais isoler le problème akratique au niveau de la connexion du résultat de la délibération et de la décision peut sembler réducteur dans la mesure où nombre de cas intuitifs d'épisodes de faiblesse de volonté n'impliquent pas ce genre de problèmes.

On peut, à la suite d'Amélie Rorty (1980), identifier, au cours de toutes les étapes du processus qui mène à la formation du meilleur jugement des défaillances et points de rupture qui seraient responsables d'épisodes de faiblesse de volonté. Par exemple, au niveau de la génération des alternatives, un agent amorphe, acidique ou apathique échouera à enclencher une recherche d'options plus satisfaisantes. Avachi sur le sofa, je peux, par exemple, constater qu'il n'y a rien de bon à la télévision, mais tout de même m'ennuyer à la regarder parce que je ne me demande même pas ce que je pourrais faire d'autre.

D'ailleurs, une hyper génération d'alternatives peut avoir un effet contraire d'hyperactivité, mais dilue pour ainsi dire les objectifs. Une personne hyperactive s'engage spontanément dans toutes sortes d'activités qui ne sont cependant pas orientées du début à la fin par des objectifs ou des buts stables. Il se peut que l'hyper génération d'alternatives surcharge en aval le reste du processus délibératif comme l'évaluation et la sélection des options (Kalis et coll., 2008). Mais peu importe l'explication, certains cas putatifs d'akrasie relèveraient en fait de ce genre de difficulté procédurale. Un agent peut se montrer trop indécis ou apathique parce qu'il génère mentalement trop d'options à examiner ou être incapable de se montrer résolu parce que les alternatives qu'il choisit manquent d'orientations et sont sujettes à une constante révision.

L'akrasie peut également résulter d'un problème situé non pas au niveau de la génération des alternatives, mais au niveau de leur sélection. Un agent intoxiqué par l'alcool ou sexuellement excité vit notamment une modification dans son aversion au risque et est, par conséquent, plus disposé à accomplir des actions qu'il regrettera par la suite. L'impulsivité

et la compulsivité relèveraient typiquement d'un problème de sélection des alternatives. Ce serait également le cas de l'indécision découlant d'une ambivalence – l'ambivalence étant la difficulté de trancher entre deux options également attrayantes, mais pour des raisons différences.

Mele appelle « *last-ditch akrasia* » certaines difficultés à exécuter une action post-décisionnelle. Cela peut relever d'un problème moteur (sur-inhibition ou sous-inhibition) ou d'une difficulté à maintenir suffisamment son attention sur l'objectif et ses étapes de réalisation. On a affaire alors à des formes d'apathie ou de suractivité motrice.

Tous les problèmes mentionnés peuvent être à la source de cas d'akrasie. Par fiat décisionnel, on peut par contre stipuler que ces cas ne relèvent pas de la stricte akrasie, mais sont des instances de faiblesse de volonté dans un sens plus général. Dans ce cas, le débat internaliste/externaliste devient par contre moins central dans le traitement théorique du problème de la faiblesse de volonté.

6.2.3 Quelques problèmes avec la notion de jugement inconditionnel de Davidson

Comme il a donné lieu à une panoplie de commentaires et de critiques depuis sa publication, on peut dire que l'article de Davidson est une référence pour n'importe quel philosophe qui aborde le problème de la faiblesse de volonté. Mais, malgré ses qualités, je ne suis pas sûr qu'il présente le même niveau d'intérêt pour les psychologues et les théoriciens de la décision. Je pense qu'on peut adresser à Davidson un certain nombre de critiques et de remarques qui relativisent considérablement sa contribution.

Il semble en premier lieu que la solution de Davidson laisse complètement intact le problème de la stricte akrasie. Comme l'on fait remarquer Bratman (1979) et Mele (1987), beaucoup de cas de faiblesse de volonté impliquent des agents qui, semble-t-il, décident à l'encontre non seulement de leur meilleur jugement, mais de leur jugement préférentiel inconditionnel. Par exemple, les fumeurs qui souhaitent se départir de leur mauvaise habitude sans en être capables n'estiment pas seulement qu'il est préférable toutes choses considérées de ne pas fumer, mais qu'il est préférable tout court de ne pas fumer. Ce point ne doit pas faire l'objet d'une critique si cette critique présuppose, comme elle semble le faire, que la stricte akrasie est un phénomène réel, et non pas l'objet du litige. Nous avons des intuitions vagues selon lesquelles la stricte akrasie existerait. Mais nous ne pouvons pas

faire une critique substantielle de la position de Davidson sous le prétexte que la conception qu'il défend de la nature de la délibération est incompatible avec ces intuitions, car nous risquerions alors de nous voir accuser de faire une pétition de principe.

Par contre, nous avons des raisons de penser que la notion de meilleur jugement inconditionnel est plus un artifice théorique que le résultat d'une conclusion naturelle tirée de l'observation des étapes du processus décisionnel réel des agents. Bien que la théorie de Davidson soit élégante, elle semble davantage reposer sur une analyse de la structure grammaticale et des propriétés logiques des phrases et énoncés censés *exprimer ou décrire* les étapes de la délibération que sur la délibération elle-même. Aussi, j'exposerai dans les prochaines sous-sections des critiques que suscite l'introduction de la notion de meilleur jugement inconditionnel et de son statut théorique. Non seulement la distinction entre jugements conditionnel et inconditionnel est précaire, mais il s'agit d'une étape cognitive superflue. D'ailleurs, je ne crois pas qu'un mouvement théorique qui assimilerait les jugements inconditionnels à des actions ou à des décisions pour les rendre moins suspects[65] serait satisfaisant parce qu'il ne permettrait pas de solutionner le problème de la stricte akrasie.

6.2.3.1 Jugements inconditionnels et conditionnels : une distinction précaire

Il n'est pas clair que la distinction entre jugements conditionnel et inconditionnel a les reins suffisamment solides pour faire le travail conceptuel que Davidson veut lui faire faire. Davidson tient pour acquis que la clause « *toutes choses considérées* » attribue une propriété intrinsèque à un jugement. Or, on peut tout à fait interpréter cette clause comme un simple indicateur de la manière dont l'agent a établi son jugement préférentiel. En affirmant que je préfère, toutes choses considérées, *a* à *b*, cela peut vouloir dire de manière elliptique que je préfère *a* à *b parce que* j'ai considéré toutes choses ou que je préfère *a* à *b et* que cela a été établi en considérant toutes choses. Je pense que ces interprétations sont intuitivement au moins aussi valables que celle de Davidson. Aussi, il est difficile d'exclure l'hypothèse que l'expression propositionnelle d'un jugement préférentiel inconditionnel putatif puisse être en fait une manière elliptique d'exprimer des jugements *toutes-choses-considérées*, et même dans certains cas des jugements *prima facie*. En fait, l'énoncé :

[65] Ce que Davidson refuserait sans doute de faire.

« *l'agent A préfère* x *à* y » est passablement ambigu et peut autoriser plusieurs interprétations. Il peut décrire un ordonnancement appréciatif à n'importe quelle phase de la délibération rationnelle. Il peut décrire un résultat temporaire aussi bien que final. Je peux examiner, par exemple, au cours d'une soirée des prospectus proposant des destinations touristiques en compagnie de ma femme et lui demander à la fin ce qu'elle préfère. Elle pourrait me répondre du tac au tac qu'elle préfère Paris à Rome. S'agit-il d'un jugement inconditionnel, *prima facie* ou *toutes-choses-considérées*? La structure logique putative du jugement n'est d'aucun secours dans ce genre de situation. Tout ce qu'on peut dire est que, *jusque-là*, ma femme préfère aller à Paris qu'à Rome, et si la décision lui revenait et qu'elle devait la prendre maintenant, *elle choisirait* d'aller à Paris.

Je pense qu'on peut critiquer Davidson d'avoir confondu les propriétés logiques des énoncés qui expriment des préférences avec les propriétés des préférences. Il est plus judicieux de concevoir les jugements préférentiels comme des ordonnancements appréciatifs (prospectifs ou rétrospectifs) d'éléments virtuels ou réels qui ponctuent les phases de la délibération et dont les propriétés intrinsèques ne varient pas en fonction de la manière dont on les décrit. En ajoutant les clauses « prima facie », « toutes-choses-considérées » et même « tout court » à notre description, on ne fait qu'indiquer des propriétés extrinsèques, comme la manière dont l'agent est arrivé à cette conclusion pratique, ou à quel résultat d'étape il en est rendu dans sa délibération. Ou alors on indique qu'on n'a pas cette information ou même que l'agent n'a pas cette information. Dans ce dernier cas, je peux expliquer, par exemple, pourquoi Bill s'est comporté de telle façon en disant que c'est parce que Bill préfère *a* à *b* tout court, ignorant les raisons expliquant la préférence de Bill ou sachant que Bill lui-même ignore, *au moment où il prend sa décision*, les raisons qui l'on amené à établir cet ordonnancement dans le passé, ni même s'il s'agit de raisons complètes (auquel cas sa préférence est un jugement *toutes-choses-considérées*) ou partielles (auquel cas sa préférence est un jugement *prima facie*).

Une autre critique qu'on peut faire de la distinction de Davidson est que nous n'avons pas l'impression qu'un jugement préférentiel inconditionnel est formé pour aucune raison, mais seulement comme l'effet d'une cause. Or, à moins d'être sous l'emprise d'une pensée magique et de se mentir à soi-même, l'akratès peut reconnaître le fait qu'il décide d'agir *pour de mauvaises raisons*, et non en raison de causes étrangères à son système mental

conscient – comme c'est le cas pour les personnes qui entretiennent des croyances motivées. Mais si Davidson a raison, alors nous ne pouvons pas demander à quelqu'un pourquoi il préfère une tarte au sucre à un bol de fruits *tout court* en espérant obtenir une réponse, parce que la réponse qu'il nous donnera nous fera tout simplement changer de sujet. Si la personne soutient qu'elle préfère la tarte parce qu'elle produira chez lui un plus vif plaisir que de manger un bol de fruits, c'est que, *prima facie*, la tarte au sucre est plus avantageuse en raison de la vivacité du plaisir qu'elle donne. En fait, on ne peut pas expliquer notre jugement préférentiel inconditionnel à l'aide de raisons, ni dans les cas d'akrasie – comme Davidson le soutient d'ailleurs – *ni dans tous les cas de comportements rationnels*.

Cette dernière implication, je pense, campe subtilement Davidson dans le camp des sceptiques humiens à l'égard du caractère cognitif des préférences. Il y aurait quelque chose d'*irréductiblement* non *rationnel* (à distinguer de irrationnel) dans la motivation humaine et qui se trouverait cristallisé ou précipité dans les jugements préférentiels inconditionnels[66]. Des forces plus ou moins obscures s'immisceraient dans la motivation par le biais – il s'agit ici d'une interprétation libre du propos de Davidson – du jugement préférentiel inconditionnel. Le terme «s'immiscer » n'est même pas adéquat parce qu'il suggère que ces forces ne seraient qu'*à l'occasion* présentes dans la motivation, alors qu'elles y seraient *toujours* présentes. Ces forces motiveraient dans toute situation l'agent sans que celui-ci puisse considérer qu'il s'agit de raisons qu'il pourrait soumettre à la critique. Ces forces irrationnelles nous pousseraient souvent à prendre de bonnes décisions, d'autre fois non. Les décisions akratiques et l'adhésion à des croyances motivées appartiendraient à cette dernière catégorie.

Il y aurait évidemment beaucoup à dire sur cette forme de scepticisme humien, et l'on pourrait sans doute lui consacrer une thèse complète. Mon objectif ici n'est pas de mentionner des arguments que je juge suffisamment convaincants pour rejeter la distinction

[66] Dans sa version plus classique, le scepticisme humien consiste à nier que le *contenu* de nos désirs, de nos aversions, de nos souhaits, de nos envies, de nos goûts, de nos penchants, de nos sentiments, de nos émotions, de nos « feelings », etc. ait un caractère rationnel ou irrationnel. D'où, l'injonction que les goûts ne se discutent pas. Mais on peut rejeter la thèse sceptique à l'égard du contenu de ces *appréciations*, tout en continuant à défendre la thèse qu'il y a quelque chose d'irréductiblement non rationnel dans la motivation d'un organisme à faire quelque chose. À la suite de Heath (2008 : 134-137), on peut qualifier cette position de version extrême du scepticisme humien.

de Davidson, mais seulement d'identifier un aspect qui est beaucoup plus controversé qu'il ne le semble à première vue.

Du reste, il n'est pas dénué d'intérêt de mentionner ici une nouvelle orientation donnée à la théorie de la décision en économie comportementale et en psychologie du choix, à savoir l'étude du phénomène du *mal-vouloir* (*miswanting*). Des chercheurs de premier plan comme Daniel Kahneman (2006), Timothy T. Wilson et Daniel T. Gilbert (2000) ont mis en relief le caractère cognitif de la plupart des volitions en les assimilant à des anticipations ou prédictions d'état affectif futur, et plus précisément d'états hédoniques. En observant les enfants en bas âge acquérir la maîtrise de mots comme « *vouloir* » et « *aimer* », on se rend compte, affirment Wilson et Gilbert, qu'ils ont appris ces mots très tôt dans leur développement linguistique, et en tandem. Cela laisse croire que ces mots décrivent des processus complémentaires, et dont l'un de ces processus, la formation de la volonté, a un caractère cognitif indubitable. En fait, beaucoup d'expériences de laboratoire, et dans des environnements moins artificiels, tendent à mettre en relief le fait que les agents anticipent piètrement leurs états affectifs futurs, notamment dans des situations d'incertitude. Ces recherches sont évidemment hautement pertinentes pour raffiner notre conception de la faiblesse de volonté, et j'en discuterai dans le chapitre portant sur la cause de l'akrasie.

6.2.3.2 Former un jugement inconditionnel : une étape cognitive superflue

Une autre question qu'on peut se poser est pourquoi il est nécessaire pour les agents de former un jugement inconditionnel pour s'engager dans une ligne de conduite? À moins d'identifier ces jugements avec des actions ou des décisions (ou intentions d'agir), on ne voit pas en quoi il serait pertinent ou important pour l'agent de former une couche supplémentaire de préférence, qui aurait les propriétés putatives des jugements inconditionnels, avant de faire un choix. Une fois qu'un agent a, au moyen d'une délibération rationnelle, établi ce qui est préférable de faire à la lumière de tous les faits qu'il estime pertinents, il peut, semble-t-il, prendre une décision motivée par cette préférence. En fait, on ne voit pas comment on pourrait tester empiriquement l'hypothèse que la couche de jugements préférentiels inconditionnels constitue un engrenage essentiel à la prise de décision. Si l'on demandait à des personnes de prendre une décision aussitôt qu'ils ont formé un jugement préférentiel conditionnel et à d'autres de le faire seulement après avoir formé un jugement inconditionnel, observerait-on des différences? Les premiers

seraient-ils plongés dans l'indécision? On peut en douter. La délibération rationnelle, pour reprendre Rawls (1973), est une activité comme une autre et le fait de savoir jusqu'où s'y engager peut faire l'objet d'une décision rationnelle. Si je gagne du temps en sautant l'étape de la formation d'un jugement préférentiel inconditionnel et que cela ne m'expose pas de manière significative à l'échec, alors je dois m'abstenir de franchir cette étape.

On peut répondre à cette critique en stipulant que les jugements inconditionnels, tels que Davidson les voit, ne sont pas formés dans une phase particulière de la délibération, par synthèses successives de préférences de niveau inférieur, mais qu'ils reflètent en quelque sorte, ou résument dans une concision maximale, le processus en entier. Un tel jugement aurait un caractère holiste, alors qu'un jugement préférentiel conditionnel aurait un caractère analytique. On pourrait toujours décomposer les préférences établies sur des raisons en d'autres préférences jusqu'à ce qu'on atteigne le « bottom-rock » des préférences fondamentales. Tandis que les préférences inconditionnelles ne pourraient pas être décomposées de la sorte. Si ma lecture de Davidson est correcte, il avait probablement une hypothèse de ce genre en tête lorsqu'il affirma dans une formule vague que « […] *tout jugement est effectué à la lumière de toutes les raisons en ce sens qu'on l'effectue en présence de, et qu'il est conditionné par, cette totalité.* » (1970 : 63) D'ailleurs, ce serait ce reflet ou résumé ou appréciation holistique qui permettrait à l'agent de sortir de l'indécision et d'enclencher une série d'actions – d'où son caractère nécessaire et l'impossibilité d'agir intentionnellement à l'encontre.

Or, je pense que si l'on considère que les jugements préférentiels inconditionnels ne sont pas des résultats d'étape dans un processus de traitement de l'information en série comme la délibération, alors c'est qu'on défend une position qui s'apparente à une version (faible) de la théorie des préférences révélées, mais à cette différence près que ce sont seulement certaines préférences – les jugements inconditionnels – qui sont à la « surface » des décisions ou intention d'agir et non toutes les préférences de l'agent. Or, comme pour toute préférence conçue comme étant à la surface des décisions, les jugements inconditionnels ne confèreraient aucun caractère prédictif et non trivialement vrai à la théorie qui les utiliserait.

Pour ce qui est de l'hypothèse de la nécessité du caractère holiste du jugement inconditionnel, on aurait de bonnes raisons de la trouver insatisfaisante. Peut-être existe-t-il

de tels jugements, mais il est plutôt évident qu'on peut s'en passer pour décrire ce qui se passe lorsqu'un agent tente de résoudre certains problèmes cognitifs. Par exemple, pour additionner de grands nombres, nous devons procéder par étapes et nous concluons par la dernière étape en espérant que nous n'avons commis d'erreurs à aucune d'entre elles. Or, en concluant, nous avons notre réponse et n'avons pas besoin d'apprécier l'ensemble du processus pour l'apposer ; nous pouvons l'apposer comme un simple résultat d'étape.

6.2.3.3 Assimiler les jugements inconditionnels à des actions : laisser le problème intact

Dans sa volonté de distinguer les deux types de jugement sur des bases autres que celle de leur structure logique, Davidson semble suggérer, dans un autre article, qu'un jugement inconditionnel est si intimement connecté à l'action qu'il se confond même avec elle :

> « *On ne peut associer directement les jugements prima facie à des actions, car il n'est pas raisonnable d'accomplir une action parce qu'elle a une caractéristique désirable. Le fait que l'on croie qu'une action a une certaine caractéristique désirable constitue une raison d'agir, mais* le fait que l'action soit accomplie constitue un autre jugement*, selon lequel la caractéristique désirable était suffisante pour nous pousser à agir – c'est-à-dire que d'autres considérations ne l'ont pas emporté sur cette caractéristique.* Le jugement qui correspond à, ou peut-être qui est identique à, l'action, ne peut donc pas être un jugement prima facie*; ce doit être un jugement catégorique ou inconditionnel qui, si nous devrions l'exprimer en mots, aurait la forme suivante : "Cette action est désirable". »* (1978 : 140)[67]

On voit que Davidson ici oscille entre l'idée de traiter les jugements préférentiels inconditionnels comme des actions (ou des attributs d'action) et l'idée de les traiter comme des ordonnancements d'actions. Si les jugements inconditionnels sont des actions, alors il n'a pas expliqué comment l'akrasie est possible, si ce n'est qu'il y a un fossé logique entre les jugements *toutes-choses-considérées* et les actions.

6.2.3.4 Assimiler les jugements inconditionnels à des décisions : le spectre de la théorie des préférences révélées

Mais on peut douter de l'hypothèse que Davidson ait vraiment eu l'intention d'assimiler les jugements inconditionnels à des actions même si certains passages de son œuvre le suggèrent. Aussi, nous avons certaines intuitions selon lesquelles que les jugements inconditionnels, tels que décrits pas Davidson, se situeraient en quelque sorte *entre* les jugements *prima facie* et les actions. Par exemple, un mécanicien peut soutenir que *tous les*

[67] Souligné par moi.

indices qu'il a observés *militent en faveur* de l'hypothèse que le système d'injection est défaillant et non le système de prise d'air, sans pour autant *conclure* que c'est le système d'injection qui est défaillant, et *tenter* ensuite de le réparer. Seulement, il y a tout lieu de penser que, dans cette situation, conclure qu'une hypothèse est préférable *tout court* est moins une préférence qu'une décision ou un choix à l'égard de l'hypothèse à privilégier. Cela revient tout simplement à dire que, après un examen des options en lice, j'en choisis une. Et il n'est pas contre-intuitif de dire que les décisions des agents peuvent autant porter sur le choix d'une hypothèse abstraite que sur des actions concrètes : on peut décider d'accomplir une action, mais également d'adhérer à une opinion, une croyance, une cause, une hypothèse, une idée, une valeur, une théorie, un principe, etc.

Or, si décider ou choisir semble correspondre d'une certaine manière au jugement inconditionnel, alors on se retrouve de nouveau à la case départ. En effet, chez un agent strictement akratique, il y aurait justement un problème de connexion entre son meilleur jugement et la décision de s'engager dans une ligne de conduite particulière.

D'ailleurs, si l'on persiste à assimiler les jugements inconditionnels à des décisions, alors cela nous oblige à adhérer proposition centrale de la théorie des préférences révélées (PPR). On peut rejeter cette implication parce qu'on estime que la théorie des préférences révélées repose sur un souci méthodologique et des présupposés épistémologiques douteux. Mais si l'on situe les jugements inconditionnels *à la surface* des décisions des agents, alors on peut se passer des autres types d'indices pour ces préférences, comme un partisan de la théorie des préférences révélées. Comme je l'ai, entre autres, mentionné dans ma critique de la proposition (PPR), situer les préférences trop près de la surface des décisions et choix effectifs, a pour effet de rendre nos attributions de préférences trivialement vraies, mais non informatives et sans caractère prédictifs.

6.3 Vouloir, évaluer et apprécier : les rapports introspectifs et le mécanisme neurologique de la sensibilisation motivationnelle

Le débat internalisme/externalisme sur la possibilité de la stricte akrasie soulève deux questions très générales au sujet de la motivation. La première question concerne la fiabilité des rapports introspectifs portant sur les motivations réelles des agents. Si j'affirme qu'il est préférable, toutes choses considérées, que je choisisse *x*, est-ce que je peux me tromper sur

mes motivations réelles même si je suis capable d'énumérer des motifs. La seconde question concerne la possibilité du découplage entre les évaluations des alternatives et la motivation à les choisir. En d'autres termes, si j'estime qu'il est préférable que je choisisse *x*, est-ce que cela implique que j'aurais une motivation à le choisir?

La réponse à la première question est visiblement positive. Les agents peuvent en dépit de leur sincérité se tromper sur leurs motivations réelles. Ils peuvent rapporter avoir consulté le menu interne de leurs préférences ou avoir délibérément formé des jugements dans leur quête d'une solution à un problème décisionnel, mais rester aveugles à leurs véritables motifs pour ces choix. Depuis la publication du célèbre article de Nisbett et Wilson, « Telling More Than We Can Know : Verbal Reports on Mental Processes » (1977), on mesure un peu mieux le degré de fiabilité de certains types de rapports introspectifs. Les chercheurs ont mis en relief le fait que les agents humains ont toujours une explication pour leurs décisions, évaluation et réaction émotives, mais pas pour des processus perceptifs ou mnésiques. Les personnes avouent d'emblée ne pas savoir, par exemple, pourquoi ils sont victimes de telle ou telle illusion d'optique ou pourquoi ils se souviennent si bien dans certaines occasions de choses peu importantes et moins des choses plus importantes. En revanche, les personnes ont toujours une réponse à donner lorsqu'on les questionne sur un choix ou une évaluation qu'ils ont réalisée. Pourtant, les personnes sont susceptibles de tromper tout autant sur leurs processus décisionnels et évaluatifs. Les résultats expérimentaux qui ont obtenu montrent que les personnes n'ont pas un accès privilégié, via l'introspection, aux processus qui lient les stimuli aux réponses qu'ils induisent. Lorsque les stimuli sont des causes *non plausibles* aux yeux de la personne, celle-ci a tendance à l'exclure des explications potentielles ou même, dans certaines situations, à ignorer son existence. En fait, les personnes consultent moins leur mémoire, lorsqu'on le demande d'expliquer une de leurs réactions, qu'ils ne forment un jugement de plausibilité sur les causes probables de leur réaction. Si l'on demande à une personne d'expliquer pourquoi elle a été aussi contente de se rendre à une fête particulière, et qu'elle répond que c'est parce qu'elle apprécie les gens qui s'y trouvent, il est probable que la personne exprime en fait une théorie sur les raisons pour lesquelles les gens aiment, en général, se rendre à des fêtes.

Les théories que les personnes utilisent, la plupart du temps de manière implicite, reposeraient sur des règles de jugement portant sur la probabilité ou fréquence de

réalisation d'événements, sur la qualité ou la représentativité des échantillons, sur des caractéristiques générales de la causalité et sur des stéréotypes culturels. Tous les facteurs qui sont susceptibles d'altérer l'un ou l'autre de ces paramètres sont, par conséquent, susceptibles de modifier les rapports introspectifs. Par exemple, la plupart des personnes adhèrent implicitement à l'idée que les causes doivent être proportionnelles aux effets[68]. Elles filtreront certaines causes possibles de leurs comportements dans l'éventualité où celles-ci ne sont pas proportionnelles à ceux-ci. Aussi, les insomniaques attribuent spontanément leur difficulté à dormir à des sources de stress importantes ou à leurs conditions actuelles de vie et non à des événements anodins comme des activités excitantes avant d'aller au lit, les heures de couché irrégulières ou la température trop élevée de leur chambre à coucher, etc.

Pour Nisbett et Wilson, les personnes ont de manière générale tendance à ignorer l'impact causal d'éléments contextuels, d'éléments absents, des comportements non verbaux, et de mécanismes subtils du jugement (ex. : l'ancrage, l'effet de halo, etc.). Pour des raisons psychologiques, mais aussi culturelles, leur attention est davantage orientée vers d'autres types de causes *plus saillantes*.

Nisbett et Wilson mentionnent tout de même que s'il existe de bonnes raisons de croire que les personnes n'ont pas un accès privilégié à leurs *processus* mentaux, ils sont tout de même un accès privilégié, mais pas automatiquement infaillible, aux *résultats* de ces processus – et dans certains cas à des résultats d'étape. Donc, le rapport introspectif des évaluations et motivations les plus fortes que les personnes disent avoir, peuvent s'avérer être une source fiable pour établir leurs réelles évaluations et motivations. Il est seulement hasardeux de se fier à ces rapports pour décrire le processus réel qui a mené à la formation de ces évaluations et motivations. Cela laisse le problème de la stricte akrasie relativement intact.

Toutefois, Lamb et coll. (1991) et Fischman et coll. (1992) ont élaboré des expériences neurologiques qui mettent en relief des désirs et appréciations qui passent complètement sous le radar de la conscience de consommateurs de drogue qui rapportent avoir d'autres désirs et expériences hédoniques.

[68] C'est ce qui expliquerait d'ailleurs pourquoi les théories du complot sont si populaires.

Lamb et coll. ont offert à des héroïnomanes l'occasion d'abaisser des leviers reliés à un dispositif censé leur administrer soit une dose d'héroïne, soit une solution d'eau saline. Après avoir actionné à plusieurs reprises les dispositifs, les sujets devaient établir un verdict d'appréciation des doses administrées. Sans surprise, les sujets ont affirmé n'avoir pas senti d'effets hédoniques positifs découlant de l'activation du premier dispositif (l'eau saline) et que, par conséquent, ils ne seraient pas disposés à essayer de s'en procurer sur le marché noir. Et sans surprise, ils ont également affirmé avoir éprouvé une sensation plaisante à la suite de l'activation du second dispositif et qu'ils seraient par conséquent disposés à employer des ressources importantes pour s'en procurer. Le plus intéressant, est que les sujets ont évalué consciemment des toutes petites doses d'héroïne de la même façon que les doses d'eau saline. Les chercheurs ont ainsi observé que les sujets ont, en fin de compte, activé autant de fois le dispositif leur administrant des petites doses que celui leur administrant de plus fortes doses. Pour les chercheurs, la fréquence des auto-administrations de drogue est un meilleur indicateur de leur volonté et de leur réelle appréciation que les rapports conscients.

Fischman et coll. ont élaboré une variante de cette expérience avec des cocaïnomanes auxquels on administrait du desipramine, une substance qui réduit considérablement la sensation subjective de manque. Ils ont établi que même si les sujets affirment ne pas vouloir obtenir une dose de drogue, et même si la desipramine produit chez eux une sensation déplaisante lorsqu'ils s'en administrent une, cela ne change rien à la fréquence à laquelle ils continuent de s'en administrer. La desipramine aurait, entre autres, pour effet de bloquer certaines voies de communication à l'intérieur du cerveau nécessaire à la computation consciente de volitions et d'appréciations.

Pour Berridge et coll. (Berridge & Winkielman, 2003 ; Berridge & Robinson, 2003), la non-conscience des motifs réels de choix est analogue au phénomène étrange de la vision aveugle (*blindsight*). Les patients atteints de ce syndrome sont capables d'identifier dans une certaine mesure des objets et de décrire leurs formes et positions spatiales sans qu'ils soient conscients de ce qu'ils voient. Lorsqu'on leur demande de situer dans l'espace un objet, ils répondent spontanément qu'ils en sont incapables, étant donné leur état de cécité, mais arrivent tout de même à le faire avec assez de succès lorsqu'on leur demande de procéder de manière aléatoire.

La réponse à la seconde question mentionnée plus haut est un peu plus difficile parce qu'elle dépend de la manière dont on interprète le terme « évaluer ». S'il s'agit d'une forme ou d'une autre d'appréciation subjective (plaisir, sentiment, sensation positive, etc.) *actuelle* découlant de la consommation de certains biens, alors on a des raisons de croire que l'évaluation et la motivation ne sont pas les deux faces d'une même pièce. Mais si par évaluation on entend un processus conscient ou inconscient d'*anticipation* d'un état hédonique futur, il est plus difficile de trancher. Examinons d'abord pourquoi on a des raisons de penser que l'évaluation est une appréciation subjective et peut être un processus découplé de celui qui est responsable de la motivation.

Berridge et coll. ne se sont pas contentés d'affirmer et de défendre l'hypothèse que les agents peuvent avoir à l'occasion des désirs inconscients et procéder à des appréciations inconscientes. Ils ont mis également en relief le fait que, dans certains cas, l'appréciation (*liking*) même inconsciente découlant de la consommation de certaines substances n'affecte pas la motivation (*wanting*) des toxicomanes à obtenir ou éviter celles-ci. En fait, il y aurait deux substrats neuronaux distincts des phénomènes de l'appréciation et de la motivation (Berridge, Robinson : 2003 ; Berridge, Robinson : 2008). Pour Berridge et Robinson, « […] *le neurotransmetteur qui fut traditionnellement le plus mesuré pour évaluer le plaisir sensoriel, la dopamine, s'avère être ni nécessaire ni suffisant pour générer une appréciation* » (2003 : 509)

L'injection d'une substance antagoniste au système dopaminergique a permis aux chercheurs d'isoler les substrats neuronaux responsables de la motivation et de l'appréciation hédonique. Les agents dont le système dopaminergique est inhibé demeurent capables d'apprécier des récompenses comme des cigarettes, des drogues ou des aliments sucrés, sans toutefois manifester la moindre disposition à essayer de les obtenir. Cette découverte a donné un éclairage nouveau au phénomène de la dépendance. L'explication standard du comportement du toxicomane est qu'il cherche à obtenir du plaisir en se droguant ou cherche à éviter du déplaisir lié au manque. Or, ce genre d'explication est pour Berridge incomplet. Beaucoup de toxicomanes cherchent à obtenir une dose même s'ils savent que la quantité disponible n'est pas suffisante pour leur procurer du plaisir. De même, un toxicomane peut avoir envie de consommer alors qu'il ne vit pas un épisode de manque. Cette envie devient même exacerbée lorsqu'il vient tout juste de consommer.

Aussi, même après une période de sevrage prolongée et quand les effets du manque se sont dissipés, beaucoup de toxicomanes retournent à leurs anciennes habitudes.

Pour expliquer ces phénomènes, Berridge développe l'hypothèse d'un problème d'alignement entre l'appréciation et la motivation dans la dépendance. Le système dopaminergique serait activé par des signaux externes qui possèdent, pour l'agent, une *saillance motivationnelle* acquise par un processus de *sensibilisation neuronale*. La sensibilisation (*sensitization*) neuronale à l'égard de certains signaux serait le fruit de processus d'apprentissage *associatifs* de type stimulus-stimulus, stimulus-réponse ou stimulus-renforcement, ou *cognitifs*, comme les attentes de récompense ou la compréhension d'un lien causal impliquant une récompense (Berridge & Robinson, 1995, 2003). Aussi, pour Berridge et Robinson,

> « *Les dépendances seraient dues au vouloir sensibilisé – pas à l'appréciation. Normalement, apprécier – le plaisir engendré par la rencontre d'une nouvelle incitation – sert d'amorce pour activer et diriger le vouloir (via l'apprentissage associatif). La force d'un vouloir à l'égard d'une incitation ordinaire dépend essentiellement du degré d'appréciation. Mais chez les toxicomanes, en raison de la sensibilisation (*sensitization*) de leur vouloir, ces deux processus en viennent à être découplés.* » (1995 : 73-74)

Si l'on se fie à ce que les recherches dans ce domaine suggèrent, la manière dont fonctionne le système dopaminergique fait de la motivation quelque chose qui ne consiste pas en premier lieu à pousser l'agent dans une recherche de plaisir, mais plutôt dans une recherche de ressources. Cela correspond d'ailleurs aux comportements d'achat de certains consommateurs qui recherchent avant tout de « bonnes affaires » sans prendre le temps d'évaluer leur besoin. Ces consommateurs se retrouvent avec toutes sortes de biens dont ils ne tirent qu'une jouissance limitée. La prise de conscience de ce fait n'est souvent pas suffisante pour les orienter dans des habitudes de consommation plus adéquates. À cet égard, le cerveau de ces consommateurs compulsifs ressemble à ceux des toxicomanes décrits par Berridge.

Cela dit, l'un des faits les plus troublants que Berridge et coll. ont découverts est que le processus de sensibilisation accrue à des signaux particuliers ne s'accompagne pas d'une appréciation accrue au point de consommation, c'est plutôt l'inverse qui se produit. Plus un signal acquiert une saillance motivationnelle pour un agent, moins la substance, le bien ou l'activité, qui lui sont causalement associés, auront un impact hédonique positif pour lui. Par exemple, il y a des jeux vidéos qui se distinguent des autres par le fort degré

d'accoutumance qu'ils produisent. Ces jeux procurent initialement beaucoup de plaisir, mais offrent un rendement marginal décroissant important qui ne s'accompagne pas d'une baisse de son attrait pour son consommateur. En fait, le consommateur a tendance à jouer de plus en plus en dépit du fait que son expérience hédonique s'appauvrit.

Tout ce qui précède milite manifestement en faveur d'une position comme celle de Mele. Les agents peuvent se méprendre sur leurs propres motivations même si on leur a demandé de former un meilleur jugement. Toutefois, rien ne nous empêche d'identifier des facteurs de choix comme la saillance motivationnelle à un critère d'évaluation relevant non pas de l'appréciation réelle du résultat des alternatives, mais de l'anticipation de ce résultat. Dans un cas de stricte akrasie, un agent pourrait inconsciemment anticiper plus de plaisir qu'il n'est disposé à le reconnaître *avant* l'acte, mais regretter amèrement son choix parce qu'il apprécie peu les résultats *pendant* et *après* l'acte. Selon Berridge et Robinson, cela est caractéristique des dépendances. Cela ne sauverait pas l'internalisme, puisque le meilleur jugement de l'agent ne dirige pas comme telle la décision (même s'il l'affecte à un certain niveau), mais plutôt une autre évaluation produite par un processus qui n'a rien à voir avec la délibération consciente.

6.4. Conclusion

Le débat sur la connexion du meilleur jugement et de la décision soulève plusieurs problèmes. La nature du jugement et ses liens logiques avec ses antécédents ou avec les causes véritables de la décision akratique, la fiabilité des rapports en première personne sur les motivations, la force causale de ces jugements, mais surtout la possibilité d'un désalignement de la motivation et des attitudes évaluatives conscientes sont au cœur du débat entre internalistes et externalistes.

J'ai exposé des raisons de penser que le meilleur jugement (sincère) garde toujours un lien causal avec la décision (hésitations, promptitudes à la révision, etc.) même dans les cas de stricte akrasie. Je ne sais pas si cela affecte en profondeur les termes du débat ou seulement de manière superficielle. On doit, par contre, tenir compte du poids causal des motifs (meilleur jugement vs les autres raisons) dans la caractérisation du débat.

J'ai également exposés des raisons de croire qu'en dépit du degré de sophistication de l'analyse de Davidson, celle-ci repose sur des distinctions conceptuelles trop artificielles et précaires pour expliquer les cas de stricte akrasie à l'intérieur d'une position internaliste. L'introduction de la notion de jugement inconditionnel comme une caractérisation du résultat de la délibération ne permet pas de formuler une solution de compromis satisfaisante entre les positions internalistes et externalistes.

J'ai exposé ensuite des raisons de croire que l'explication de la stricte akrasie par une forme de désalignement de l'évaluation consciente et de la motivation, comme le propose Mele, est une bonne avenue. Les résultats obtenus par Berridge et ses acolytes attestent que les zones neuronales responsables de la motivation sont fonctionnellement différentes des zones responsables de l'appréciation, et que les premières peuvent s'activer à l'insu des agents. Cela contredit directement la position internaliste.

Aussi, si l'on persiste à nier l'existence de la stricte akrasie, on ne peut pas le faire à partir d'une position internaliste. Le meilleur jugement peut affecter causalement le comportement de l'agent, mais il est possible que d'autres motifs (conscients ou non) aient un poids plus important et qu'ils orientent l'agent dans une autre direction. Une position sceptique est sans doute toujours possible, mais doit reposer sur d'autres bases.

Une des options envisageables pour un sceptique réside, à mon avis, du côté de l'hypothèse de la maximisation. Même si les agents peuvent décider à l'encontre de leur meilleur jugement, ils tentent néanmoins toujours d'une manière ou d'une autre de maximiser la satisfaction de leur préférence. Seulement, le processus d'évaluation des alternatives est essentiellement inconscient et peut ne pas s'enligner avec les évaluations conscientes. En d'autres mots, les agents ne peuvent pas prendre des décisions contre-préférentielles (strictement akratique), bien qu'ils puissent décider à l'encontre de leur meilleur jugement. Il s'agit évidemment d'une position sceptique beaucoup plus faible. Mais l'attention théorique se déplace dorénavant des jugements aux préférences.

D'ailleurs, si l'on accepte d'aller au-delà de la théorie des préférences révélées (comme on est obligé, je pense, de le faire), le processus de sensibilisation neuronale (*neural sensitization*), décrit par Berridge, ne poserait pas vraiment de difficultés. On doit seulement distinguer les évaluations prospectives (inconscientes) des évaluations

correspondant aux appréciations actualisées. Un agent déçu par la réalisation d'un scénario qu'il anticipait être plus positif peut être considéré tout de même comme un agent maximisateur qui décide conformément à ses préférences. Dans le cas de la sensibilisation neuronale, un agent évaluera à la hausse le degré de désirabilité d'une alternative, tout en faisant une expérience positive de plus en plus pauvre.

Cela dit, l'hypothèse de la maximisation peut recevoir une interprétation qui rend la stricte akrasie conceptuellement impossible. Même si la description empirique détaillée des processus décisionnels montre que ceux-ci n'ont rien à voir avec une forme de maximisation (due, par exemple, aux limites cognitives et aux ressources temporelles des agents), les agents se comportent « comme si » ils maximisaient la satisfaction de leurs préférences, et, à ce titre, ne peuvent pas être décrit comme des agents strictement akratiques. La maximisation « *comme si* » serait en fait une condition nécessaire à l'interprétation des décisions des agents rationnels et même de l'attribution de croyances et de désirs à ces agents. Aussi, un diagnostic de stricte akrasie violerait une sorte de vérité conceptuelle.

J'aborderai l'hypothèse de la maximisation au prochain chapitre. J'entends montrer qu'il s'agit d'une hypothèse erronée basée sur des arguments faibles, et que, par conséquent, elle ne peut pas non plus servir à justifier le scepticisme à l'égard de la stricte akrasie. Les raisons qui mènent au rejet de la maximisation nous amènent également à reconsidérer la conception Standard.

7

Quelques problèmes avec l'hypothèse de la maximisation et ses conséquences pour la conception Standard

La vie est l'art de tirer des conclusions suffisantes de prémisses insuffisantes

–Samuel Butler

7.1 Conceptions étriquées du processus décisionnel et hypothèse de la maximisation

Les conceptions du choix rationnel dans la théorie économique moderne et en philosophie reposent sur une hypothèse fondamentale, à savoir que les agents rationnels sont des maximisateurs (ou des optimisateurs). [69] Bien que nommée ainsi que depuis une période assez récente, on la trouvait déjà exprimée en filigrane dans *Le Protagoras* de Platon[70]. Cette hypothèse sert de trames de fond aux débats opposant les partisans de l'internalisme et de l'externalisme motivationnel et oriente en partie ceux-ci. En effet, la conception Standard mentionne explicitement que l'akratès décide à l'encontre de son meilleur jugement, comme si la recherche du meilleur choix était une condition de la rationalité pratique. En fait, si la faiblesse de volonté apparaît si paradoxale, ce n'est pas tant qu'elle

[69] Il y a évidemment des variations entre les théories sur la portée et l'interprétation correcte à donner à cette hypothèse. Dans les théories économiques modernes, ce qui est maximisé est la satisfaction des désirs et des préférences, peu importe le fait que cette « satisfaction » procure à l'agent (ou aux autres) une expérience hédonique positive. C'est ce que Kahneman appelle l'« utilité de décision » (*decision utility*). C'est une conception plutôt formelle et empiriquement pauvre de la notion d'utilité. Cela contraste avec les conceptions substantielles – comme celle de Bentham notamment – qui fait une large place à l'anticipation d'état hédonique futur au sein de l'espace motivationnel de l'agent. C'est ce que Kahneman appelle l'« utilité hédonique » (*hedonic utiliy*). C'est ce genre de conception qu'on retrouve actuellement dans le programme de psychologie positive initié par Selligman et Csikszentmihalyi (2000), qui est un programme de recherche sur les causes du bonheur.

Sur la portée de l'hypothèse de maximisation, certains éprouvent des réserves à l'idée d'en faire une condition de la rationalité parce que cela peut aller à l'encontre des principes moraux qui doivent orienter nos conduites. On peut considérer que les principes moraux sont des préférences pour des moyens disponibles et que l'agent qui maximise ne fait pas que considérer des conséquences, mais peut également considérer des conduites morales. Un théoricien mal à l'aise avec ce genre de proposition peut tout de même estimer que les agents rationnels maximisent à l'occasion *à l'intérieur* de contraintes morales. On peut, par exemple, me considérer comme un agent maximisateur même si je ne considère pas dans ma délibération l'option de cracher au visage d'une personne âgée pour obtenir son siège dans l'autobus, mais me contente de m'asseoir un peu plus loin. Comme agent maximisateur je vais seulement considérer les options moralement praticables et tenter d'identifier parmi celles-ci la meilleure.

[70] Voir la section 2.1.1.4.

exemplifierait un supposé manque de connexion entre les jugements et les décisions, mais qu'elle semble être un contre-exemple à l'hypothèse que les agents maximisent leur bien-être ou la satisfaction de leurs préférences. On a vu quelques propositions externalistes, mais surtout internalistes, visant à résoudre le paradoxe. Aucune de ces solutions n'est cependant entièrement satisfaisante et celui-ci semble demeurer intact.

Pourtant, il y a tout lieu de croire que ce paradoxe découle d'une hypothèse erronée. Les agents rationnels ne maximisent pas – ou du moins rarement – leur bien-être ou la satisfaction de leurs préférences. L'hypothèse de la maximisation implique que les agents délibèrent ou effectuent des calculs pour découvrir la meilleure option parce que cela est la seule manière de le faire. Or, peu de situations requièrent que nous traitions les informations de la sorte. Les agents rationnels utilisent des procédures décisionnelles variées et adaptées aux problèmes qu'ils souhaitent résoudre, et qui n'ont rien à voir avec des procédures d'optimisation. Ce fait a des conséquences importantes sur notre manière de penser non seulement la motivation en général, mais la faiblesse de volonté en particulier. Cela rend la conception Standard manifestement inadéquate. Si les agents rationnels ne maximisent que rarement, alors la conception Standard nous amène à surdiagnostiquer les épisodes de faiblesse de volonté. Dans l'éventualité où l'on souhaiterait restreindre la conception Standard à la stricte akrasie, alors non seulement cela laisserait intact le problème de faiblesse de volonté, mais cela réduirait aussi le propos à l'examen d'un phénomène beaucoup plus rare que ce qui est estimé par les théoriciens.

En exposant ces procédures et en décrivant leurs propriétés fondamentales, je pense être en mesure de montrer que le fait de décider à l'encontre de son meilleur jugement n'est pas une condition suffisante, *ni même nécessaire*, pour diagnostiquer un épisode de faiblesse de volonté. Le principe de continence que Davidson juge constitutif de la rationalité pratique est non seulement accessoire, mais inutile, voire néfaste, dans la plupart des situations de choix.

7.1.1 Maximiser au sens large ou étroit et les procédures décisionnelles les plus courantes

Pour Jon Elster, il y a un proverbe norvégien qui résume les théories du choix rationnel : « On ne traverse pas la rivière pour puiser de l'eau »[71]. Ce proverbe exprime bien, selon lui, l'hypothèse transculturelle que les agents sont des maximisateurs. Il s'agit en effet d'une hypothèse particulièrement robuste, à condition toutefois qu'on la prenne dans son *sens étroit*. Maximiser dans le sens étroit consiste tout simplement à choisir la meilleure option *quand on la connait déjà*. Aussi, on maximise spontanément dans un contexte de choix trivial où l'on n'a pas besoin d'effectuer une recherche. Or, dans un *sens large*, maximiser est plus un processus de recherche et de computation que de choix comme tel. On maximise au sens large dans des contextes de choix non triviaux où l'on doit se mettre en quête d'informations. Dans ce genre de contexte, il est toutefois très difficile et presque toujours inapproprié de maximiser. Aussi, ma critique de l'hypothèse de la maximisation porte sur son sens large et non étroit.

Dans toutes sortes de situations, les agents utilisent des procédures et des méthodes de décision variées pour faire des choix. Elles sont très différentes de celles qui sont censées déboucher sur un meilleur jugement, comme la méthode de Franklin ou la Règle de Dawes. Ces procédures et méthodes sont communément appelées « heuristiques ». Leur utilisation est prescrite par un modèle de rationalité *limitée* et s'oppose au modèle de *rationalité illimitée* qu'on retrouve dans la tradition philosophique et économique.

Bien que les agents rationnels soient réputés avoir des capacités cognitives et des ressources – principalement du temps et des énergies – limitées, les modèles de rationalité pratique que les philosophes et économistes ont élaborés au fil des siècles ne semblent pas en avoir tenu compte. Pour évaluer les décisions, la tradition philosophique, depuis Aristote, utilisait des normes qui ne s'appliquent, en fait, qu'à des agents qui ressemblent plus à des démons laplaciens qu'à des organismes pourvus d'une quantité de neurones certes grande dans certains cas, mais tout de même limitée. Les jugements pratiques corrects, comme les jugements théoriques, étaient conçus comme des preuves aussi certaines que des preuves mathématiques. On assista pendant les Lumières à une

[71] Entretien au Collège de France, La Lettre, no. 21.

modification de ce paradigme. Le fameux pari de Pascal illustrait une forme de raisonnement pouvant être jugé correct, mais qui abandonnait, du coup, la recherche de certitudes (Daston, 1988 ; Todd & Gigerenzer, 1999). Ce fut la première impulsion pour l'élaboration du calcul des probabilités. Au 20^{e} siècle, des éléments de ce type de calcul ont été intégrés aux divers modèles de rationalité pratique, comme la fameuse Théorie de l'Utilité Espérée (Newman, Morgenstein : 1944). Bien que l'abandon de la recherche de certitudes marque un pas important dans l'élaboration de modèles normatifs plus réalistes, il n'en demeure pas moins que ces modèles restaient tributaires d'une conception « démonique » de la rationalité. Mais au lieu d'avoir un démon laplacien œuvrant dans un monde entièrement déterminé – au sens métaphysique du terme – en guise de point de référence, on lui substituait un démon œuvrant dans un monde en partie indéterminé – d'où la nécessité d'inclure un calcul probabiliste dans la délibération rationnelle.

Dans les années 1950, l'économiste Herbert Simon mis en relief une procédure décisionnelle cognitivement moins exigeante, mais plus réaliste pour décrire des comportements répandus et la structure computationnelle qui les sous-tend. Ce fut le début de la recherche sur les heuristiques décisionnelles. Depuis, des chercheurs – issus surtout des départements de psychologie et d'économie comportementale – ont mis en relief toutes sortes de règles de décision qui tiennent compte de la rationalité limitée des agents et de la fugacité informationnelle de leur environnement.

J'exposerai dans cette section certaines des heuristiques décisionnelles les plus courantes et mettrai en relief leurs caractéristiques les plus importantes. J'examinerai les raisons qui motivent certains chercheurs à considérer que ces procédures et méthodes mènent, dans la plupart des cas, à des décisions rationnelles. L'objectif principal de l'exposition de ces procédures est de montrer clairement des alternatives viables à la maximisation[72].

[72] Je ne discuterai pas de l'une des conséquences importantes qui découlent du fait que les agents utilisent des heuristiques et procédures décisionnelles variées, à savoir qu'un bon modèle d'interprétation des choix ne met pas seulement en relief les désirs et les croyances des agents. Décrire l'usage de ces procédures met en relief le fait que l'explication des choix des agents – rationnels ou irrationnels – ne doit pas seulement identifier des considérations désidératives, affectives et cognitives, mais également des considérations normatives, comme des règles de choix et de procédures décisionnelles. Ces dernières considérations forment un paramètre distinct. En l'occurrence, si deux agents partagent initialement les mêmes désirs, dispositions émotionnelles et croyances, ils peuvent néanmoins être amenés à prendre des décisions différentes parce qu'ils utilisent des règles décisionnelles différences. Ce fait est, à mon sens, particulièrement significatif parce qu'il nous permet de prendre nos distances vis-à-vis du modèle classique d'interprétation des actions et des choix reposant sur les croyances et les désirs. Dans le cadre de ce modèle, les décisions akratiques sont vues comme une forme

7.1.1.1 « Satisfier »

Herbert Simon a découvert que les agents économiques, comme les marchands et les conseils d'administrations des entreprises, ne se comportent pas comme le prédisaient les modèles économiques de l'époque. Les stratégies marchandes ne sont pas adoptées à la faveur d'un calcul de risque sophistiqué qui inclurait des considérations probabilistes dans une délibération rationnelle comme la méthode de Franklin. Les agents opteraient plutôt pour une procédure consistant à fixer, *à l'avance*, un seuil de satisfaction à l'aide de critères pertinents et à sélectionner ensuite la première option qui se situe au-dessus de ce seuil. Simon a appelé cette procédure « satisfier ». Intuitivement parlant, satisfier n'est pas rechercher le meilleur, mais seulement ce qui pour nous sera satisfaisant. Par exemple, si nous cherchons une paire de chaussures, il est judicieux de se fixer à l'avance un seuil à partir des propriétés pertinentes comme le prix, l'apparence, la durabilité et le degré de confort. Nous pouvons, par exemple, stipuler qu'une paire de chaussures satisfaisante doit coûter en dessous de 100 $, être assez belle, relativement peu durable (si nous sommes prêts à faire des compromis sur ce point), mais en revanche très confortable. Une fois le seuil fixé, nous pouvons commencer la collecte d'informations. Si nous rencontrons au cours de nos déambulations dans les boutiques une paire de chaussures qui est en dessous de 100 $, et au moins assez belle, peu durable et très confortable, nous savons que nous devons la choisir.

Satisfier est une procédure conjonctive parce que le seuil préalablement fixé est constitué par la conjonction de critères dont chacun doit être satisfait par une option pour que celle-ci soit éligible. Ce qu'il faut comprendre, ici, est que, même si l'agent qui satisfie utilise les mêmes critères de choix que celui qui recherche la meilleure option ou cherche à former un meilleur jugement, il ne les utilise pas de la même façon. Si je souhaite découvrir *la* meilleure paire de chaussures, je ne dois pas me fixer un seuil à l'aide des critères pertinents pour moi, mais plutôt examiner l'ensemble des chaussures disponibles et sélectionner la paire qui maximise mes critères positifs (l'apparence, confort, durabilité) et qui minimise mes critères négatifs (prix). Si nous avons préalablement des critères de

---

d'échec (ou de réussite subtile) dans la maximisation de la satisfaction des désirs, étant données les croyances de l'agent. C'est ce qui participe, d'une certaine manière, à son caractère paradoxal. Or, l'inclusion de procédures décisionnelles et de considérations normatives variées dans notre arsenal explicatif affaiblit ce caractère.

choix, nous ne pouvons satisfier en les appliquant directement aux options examinées. Nous devons observer une étape intermédiaire, soit élaborer un seuil à partir de ces critères.

Cette procédure conjonctive est extrêmement répandue. Elle pilote beaucoup de choix, allant du choix d'un conjoint à l'achat d'une maison. D'ailleurs, elle ne concerne pas seulement le domaine des décisions pratiques, mais également théoriques. En effet, on peut choisir une hypothèse ou une avenue théorique satisfaisante sans examiner l'ensemble des avenues disponibles. Par exemple, les enquêteurs qui doivent résoudre une série de crimes graves dirigeront leurs efforts d'investigation sur le premier individu qui leur semble suspect. Ils concentreront leurs efforts dans l'examen de la première hypothèse qui leur apparaît satisfaisante avec des critères de choix, comme l'antécédent judiciaire du prévenu, son intérêt marqué dans la perpétuation du crime, le dernier contact avec les victimes, etc.

7.1.1.2 Élimination par attribut et autres méthodes lexicographiques

Un agent peut également utiliser un seuil, non pas pour sélectionner un choix, mais pour en éliminer un. À défaut d'avoir rapidement, et à peu de frais, accès à l'ensemble des informations nécessaires pour effectuer le meilleur choix, les agents peuvent éliminer les options à l'aide d'un attribut saillant. S'il reste des options en lice après, l'agent peut utiliser un autre aspect saillant et répéter les phases d'élimination jusqu'à ce qu'il ne reste qu'une option en lice, laquelle sera sélectionnée. Cette procédure décisionnelle, initialement décrite par le prix Nobel Amos Tverski (1972), est largement utilisée dans la vie de tous les jours. On s'en sert, par exemple, pour attribuer une charge d'enseignement dans une institution scolaire. Le bureau des ressources humaines opère un premier tri des candidats en éliminant les CV de candidats qui ne possèdent pas le bagage académique requis. Puis, si la quantité de candidats permet une marge de manœuvre importante, le bureau élimine ensuite les candidats qui n'ont pas d'expérience de travail dans le domaine de l'enseignement. Les autres candidats sont convoqués pour une entrevue. On élimine ensuite les candidats qui semblent ne pas s'être suffisamment préparés, puis les candidats qui ne semblent pas capter suffisamment l'attention de leur auditoire. S'il en reste plusieurs en lice après ces phases successives d'élimination, on élimine ceux qui manquent d'entregent et semblent susceptibles de s'isoler ou d'entretenir des conflits avec leurs collègues.

Dans un processus d'embauche, l'ordre des critères d'élimination n'est pas arbitraire. Il peut varier d'une institution à l'autre, mais il est censé représenter l'importance relative que les institutions attribuent à chacun des critères pour mesurer la compétence d'un travailleur. Les critères les plus importants sont habituellement les premiers utilisés pour discriminer les candidats – sauf dans les situations où il est plus commode ou plus facile d'utiliser un autre critère.

On utilise une méthode lexicographique analogue lorsqu'on évalue la pertinence ou l'acceptabilité d'une information provenant d'un expert pour s'orienter dans un domaine qu'on connaît peu. On se demande d'abord si la personne qui fournit l'information est impartiale. Si elle ne l'est pas, on rejette l'information ou refuse d'en tenir compte. Si la personne est impartiale, on doit se demander ensuite si elle est vraiment une experte. Si elle ne l'est pas, alors on rejette l'information. Si elle l'est, alors on doit se demander si elle est une experte dans le domaine qui nous intéresse ici. Si elle ne l'est pas, on rejette l'information. Si elle l'est, alors on doit ensuite se demander si l'information fournie par cet expert fait l'objet d'un consensus relatif dans la communauté des experts dans le domaine. Si ce n'est pas le cas, on doit rejeter l'information, tandis que si c'est le cas, on a alors une bonne raison de tenir compte de cette information pour prendre une décision. Il s'agit ici du schéma typique de l'appel à l'expertise tel qu'on l'enseigne dans les collèges.

Il existe également une autre procédure discriminante qui consiste, non pas à éliminer successivement des options, mais à les sélectionner sur la base d'un seul attribut à la fois. C'est la procédure appelée « *Take-the-best* » par Peter Todd, Gerd Gigerenzer et les membres de l'*ABC group* de l'Institut Max Planck. « *Take-the-best* » consiste à tenter de sélectionner l'option à choisir à l'aide de l'attribut le plus approprié (ou écologiquement valide selon ces auteurs) et de passer ensuite au second meilleur attribut jusqu'à ce qu'on réussisse à isoler une seule option, laquelle sera choisie.

Cette procédure est très utile, par exemple, dans le cas des paris sportifs. Si vous deviez parier avec un ami féru de hockey sur l'issue d'une partie, vous maximiseriez vos chances si vous utilisiez l'attribut du pointage pour sélectionner une équipe. S'il y a une différence de plus ou moins 10 points au classement en faveur d'une équipe, vous devez miser sur elle. Si ce premier attribut ne vous permet pas de départager les équipes, alors vous devez mesurer le nombre de parties gagnées et perdues dans la séquence des 10 dernières parties.

Si vous observez une différence de plus ou moins 2 points, sélectionnez l'équipe qui a fait mieux, sinon, passez au dernier attribut. Choisissez l'équipe qui joue à domicile. Des chercheurs ont montré qu'une procédure analogue offre de meilleures perspectives de gain – notamment au baseball – qu'une méthode plus lourde consistant à tenir compte d'une quantité impressionnante d'informations relatives aux joueurs et aux forces en présence.

Les publicitaires exploitent souvent le fait que les agents utilisent « *Take the best* » pour leur choix de consommation. Ils mettent en relief systématiquement un aspect positif important de leur produit et mentionnent le fait que le produit concurrent ne l'offre pas.

L'élimination par attribut et « *Take the best* » appartiennent à la catégorie des méthodes lexicographiques. Une méthode lexicographique offre toujours en principe la possibilité de trancher dès la première évaluation ou le premier examen. On s'en sert évidemment pour s'orienter dans un dictionnaire où la première lettre d'un mot examiné permet de savoir si l'on doit chercher plus loin ou revenir en arrière. Aussi, pour déterminer lequel de deux nombres constitués de plusieurs chiffres est le plus grand, on essaie de déterminer lequel en contient le plus. S'ils contiennent le même nombre, alors on passe au second attribut le plus important, soit la grandeur du premier chiffre des deux séries. Si cela ne permet pas de trancher, alors on passe au second chiffre, etc. Les méthodes lexicographiques, lorsqu'elles sont adéquates, permettent de choisir de manière fiable et avec une économie de moyen. Même si elles ressemblent à certains égards à une procédure qui mène à la formation d'un meilleur jugement, elles négligent beaucoup d'informations pertinentes (ex. : qui garde les buts; quels joueurs sont absents; etc.).

7.1.1.3 Choix par ancrage

Lorsque nous évaluons des options, nous nous servons souvent de points d'ancrage situés dans le passé. Ces ancrages servent pour ainsi dire de points de référence, et la « distance » qui les sépare des options examinées constitue l'étalon pour évaluer ces dernières. Par exemple, lors de leur premier cours, les étudiants choisissent typiquement une place en classe qu'ils choisiront par la suite chaque séance. Si, lors d'une séance ultérieure, ils voient « leur » place occupée par un autre, ils examineront spontanément une place adjacente pour s'y asseoir. On ne peut pas expliquer les comportements ultérieurs de sélection d'une place par des motivations identiques qui perdurent dans le temps. Une

quantité significative d'étudiants choisissent initialement de manière plus ou moins aléatoire. D'autres étudiants ont des motifs plus spécifiques, mais ont une préférence pour certaines zones que pour les places proprement dites. Par exemple, certains préfèrent être assis au fond de la classe parce qu'ils préfèrent regarder les autres plutôt que d'être regardés. D'autres préfèrent être près d'une fenêtre pour rêvasser plus facilement. D'autres préfèrent être plus près de la porte, histoire de s'enfuir plus rapidement à la fin du cours! On peut aussi évaluer les coiffures en fonction de la distance esthétique de celle qu'on arbore depuis des années ou même de la première coiffure qu'on a eue dans notre jeunesse (cela semble être le cas de la chanteuse française Mireille Mathieu).

L'ancrage se produit souvent – mais pas toujours – à la faveur d'une décision ou d'une expérience qu'on a prises ou eues la première fois dans un contexte donné. L'économiste Dan Ariely (2008) explique pourquoi les consommateurs ont plus de plaisir à acheter un lecteur DVD ou un bien technologique quelconque que d'acheter de l'essence ou du lait. Acheter de l'essence ou du lait comporte un élément déplaisant supplémentaire. Nous comparons spontanément le prix de ces biens avec le prix du plus lointain souvenir que nous avons du prix qu'ils avaient. Or, comme l'essence et le lait coûtent aujourd'hui beaucoup plus cher qu'à l'époque, nous en ressentons un vif déplaisir. En revanche, l'effet d'ancrage est positif lorsque nous nous procurons des biens technologiques parce qu'ils sont toujours moins chers – ceci ne vaut évidemment que pour des biens analogues ou comparables[73].

Les points de référence ne s'ancrent pas toujours dans le passé. Les agents se servent souvent d'expériences ou de décisions motivées par des attitudes égocentriques pour anticiper le comportement des autres et motiver ainsi l'usage d'une ligne de conduite stratégique de préférence à une autre, ou tout simplement pour mesurer la force d'une intuition personnelle. Thomas Schelling affirmait un jour que « You can sit in your armchair and try to predict how people will behave by asking yourself how you would behave if you had your wits about you. You get, free of charge, lots of vicarious, empirical

[73] La prévalence de l'ancrage dans les choix de consommation explique aussi pourquoi en période d'inflation les producteurs de denrées sous-dimensionnent leurs produits au lieu d'en augmenter le prix. Ils savent que les consommateurs rechigneront à acheter un bien plus cher même s'ils ont une préférence marquée pour ce bien. Il est plus facile d'établir un prix de référence par ancrage qu'une quantité de références par ancrage dans les cas où la mesure de ces quantités est subtile, plus difficile à établir parce que moins saillante.

behavior. (Schelling, 1966) »[74]. Schelling décrivait la méthodologie des chercheurs dans les sciences du comportement de l'époque. Mais cela décrit également la manière dont nous procédons souvent pour faire des prédictions concernant la conduite des autres et orienter, par conséquent, notre propre conduite.

Une autre forme de choix par ancrage est l'usage de la règle « *Take the last* ». Cette règle consiste à solutionner un problème décisionnel en optant pour la dernière solution qui a fonctionné pour un problème identique ou du moins analogue. Une procédure qui utilise *Take the last* offre des résultats satisfaisants parce qu'elle permet d'éviter des coûts de recherche prohibitifs. Or, l'agent peut tout à fait savoir que sa dernière solution n'est pas la meilleure, toutes choses considérées, mais continuer à l'appliquer chaque nouvelle occurrence du problème.

7.1.1.4 Choix par défaut et imitation

Nous imitons allègrement nos conspécifiques autant pour des décisions dont la portée est insignifiante que pour des décisions qui engagent une vie entière. L'étendue et la sophistication des comportements mimétiques des humains reposeraient sur des capacités évoluées du cerveau leur permettant d'analyser dans le détail le comportement des autres et de le reproduire fidèlement. Ces capacités distingueraient notamment les humains des autres primates. Les mécanismes qui sous-tendent l'imitation sont souvent inconscients. Parfois, il semble y avoir un voile si opaque entre la conscience des agents et l'origine mimétique de leurs décisions qu'ils refusent souvent de considérer ces dernières comme de véritables décisions. Lorsque des adolescents disent qu'ils planifient avoir des enfants lorsqu'ils seront adultes, ils refusent souvent de considérer qu'ils ont pris la décision d'avoir des enfants plus tard : ils affirment spontanément qu'ils n'ont rien décidé *parce que cela va de soi d'avoir des enfants lorsqu'on est adulte*[75]. Aussi, beaucoup se montrent surpris à l'idée de discuter de leur volonté d'avoir des enfants, mais se ravisent lorsqu'ils décident de se prêter au jeu de la critique des motifs, tantôt égoïstes, tantôt altruistes, d'avoir des enfants.

---

[74] Cité par Dawes (1990 : 180).

[75] Certains spécifient par contre des conditions plus exhaustives, comme avoir suffisamment de moyens financiers, un conjoint fiable et amoureux, une vigueur minimale, etc. Mais cela ne change rien à la nature du problème.

Il y a, d'ailleurs, toute une littérature sur le phénomène des *choix par défaut* que les agents font d'une manière plus ou moins consciente. L'une des avenues les plus prometteuses est l'étude des choix par défaut *révélés* par des modèles institutionnels. Par exemple, le choix de contribuer à un régime de retraite est, dans certains pays, un choix par défaut que les travailleurs peuvent changer à tout moment. La décision de ne pas donner ses organes après sa mort est au Québec un choix par défaut, alors que c'est l'inverse en France. Comme les banques d'organes sont beaucoup plus garnies dans les pays où le don d'organe est un choix par défaut, cela justifie ce que Thaler et Sunstein (2003) appellent un libéralisme paternaliste, soit l'attitude politique qui consiste à maintenir les libertés de choix, tout en orientant les citoyens vers des choix socialement plus judicieux en exploitant la possibilité des décisions par défaut. Ce genre de mesure institutionnelle participe de ce que ces auteurs appellent le « paternalisme libertaire ».

7.1.1.5 Élimination des extrêmes

Typiquement, une délibération est une comparaison. Mais il arrive souvent que les agents doivent faire face à une série d'options difficilement comparables avec les informations dont ils disposent. Ce n'est pas que les options comportent des aspects incommensurables, mais que certains de ces aspects sont inconnus. Et, alors, la présence des autres aspects ne permet pas d'identifier un choix qui se démarque positivement des autres en tous points. Face à un problème décisionnel de la sorte, les agents peuvent s'engager dans une coûteuse recherche d'informations, mais souvent ils refusent de s'engager dans une telle démarche et utilisent une règle simple qui consiste à éliminer les extrêmes. Par exemple, vous désirez vous acheter un téléviseur HD. Vous aimez faire de bonnes affaires, ce qui vous amène à examiner 3 téléviseurs que les marchands de votre région vendent au rabais : (1) un téléviseur Sony de 60 po à 1680 $; (2) un Panasonic de 52 po à 1320 $ et ; (3) un Samsung de 46 po à 950 $. Vous reconnaissez les trois marques, mais ne connaissez pas grand-chose aux caractéristiques d'arrière-plan des téléviseurs. Vous pouvez lire la liste de ces caractéristiques, mais vous avez de la difficulté soit à les comprendre ou à en mesurer l'importance. Pour découvrir la meilleure affaire, vous devrez vous engager dans une recherche poussée et certainement très coûteuse en temps. Refusant cette éventualité, vous allez probablement choisir le téléviseur Panasonic. Ce faisant, vous évitez le risque de payer trop cher et le risque d'être déçu par un item qui ne vous en donne pas assez.

Encore une fois, votre procédure décisionnelle n'est pas censée déboucher sur la formation d'un meilleur jugement. Vous ne savez pas s'il s'agit du meilleur choix à faire, compte tenu du fait que vous n'avez pas considéré toutes choses avant de vous décider.

7.1.1.6 Choix aléatoire

Lorsque les alternatives sont en tous points incommensurables, que notre délibération manque de fondements ou que le temps de décision est une variable importante, un choix aléatoire peut être judicieux. Les procédures de décision ne sont adéquates que dans la mesure où elles permettent de maximiser les chances de succès de nos entreprises.

Mais quand nous avons à trancher entre des options très incommensurables, il arrive que nous ayons l'impression que l'usage d'une procédure décisionnelle plus sophistiquée que le choix aléatoire ne maximise pas vraiment nos chances de succès. Si nous vivons un échec, nous pouvons nous en vouloir de ne pas nous être engagés dans une délibération plus sophistiquée. Mais les regrets *post hoc* ne sont pas nécessairement une raison suffisante pour condamner un choix. Des recherches en psychologie sur le phénomène de la dissonance cognitive tendent à montrer que les personnes regrettent davantage les choix qu'ils n'ont pas faits que les choix qu'ils ont faits. Des mécanismes de réduction de la dissonance cognitive font le travail d'atténuation des désagréments liés aux événements, aux états de choses ou aux personnes qu'on ne peut foncièrement éviter. En outre, on estime communément qu'une décision importante requiert une procédure décisionnelle sophistiquée. Or, il s'agit d'une erreur – que je vais expliquer dans la prochaine section. Un choix aléatoire peut être la meilleure procédure pour prendre une décision importante.

Lorsqu'un agent manque d'informations pertinentes pour découvrir la meilleure option, un choix aléatoire peut s'imposer. Un consommateur de billets de loterie à gratter ne délibère pas devant le commis dans le but de découvrir un billet gagnant, même si un tel billet gît probablement dans le présentoir. Comme sa délibération serait sans fondement, il procédera à un choix aléatoire.

Cela dit, la délibération demande du temps. Dans des situations d'urgence, il est quelquefois préférable d'opter pour un choix sous-optimal, et de le faire rapidement, plutôt que de rechercher le meilleur choix. Par exemple, si vous traversez la rue de manière négligente et que vous vous rendiez compte, en une fraction de seconde, qu'une voiture

fonce sur vous, vous n'allez pas délibérer pour savoir si vous devez vous jeter à gauche ou à droite de la voiture pour l'éviter. Un choix aléatoire est préférable même si un côté est plus désavantageux que l'autre – parce qu'il comporte, par exemple, un obstacle qui représente un risque de blessure.

7.1.1.7 Heuristique de la reconnaissance

Beaucoup d'organismes évolués – comme les mammifères par exemple – rechignent à ingurgiter des aliments qu'ils ne reconnaissent pas. L'heuristique de la reconnaissance est une règle simple qui consiste à effectuer une inférence à partir de la reconnaissance ou de l'absence de reconnaissance d'un objet, d'une substance, d'un attribut ou ensemble d'attributs. Il s'agit le plus souvent d'une stratégie d'évitement du risque. Nous rechignons souvent à acheter des biens dont nous ne connaissons pas la marque, même si nous pouvons, la plupart du temps, les obtenir à moindre coût. Les publicitaires savent que les consommateurs utilisent fréquemment cette procédure – ne serait-ce que pour effectuer un premier tri. Les compagnies Apple et Benetton furent sans doute les premières à exploiter cette procédure avec autant de clarté en lançant des campagnes publicitaires qui ne faisaient aucun étalage de leurs produits, mais exposaient le nom de leur marque de commerce.

Comme pour le choix aléatoire, l'heuristique de la reconnaissance appartient au lot des procédures décisionnelles les plus éloignées des méthodes d'optimisation, comme la méthode Franklin, le calcul de l'utilité espérée ou même la règle de Dawes.

7.1.2 Les caractéristiques générales des heuristiques décisionnelles

Bien que fort variées, les heuristiques comportent néanmoins des caractéristiques communes qui les distinguent clairement des délibérations complètes ou quasi complètes, comme le requiert la formation d'un meilleur jugement. Je pense qu'il est approprié d'en parler ici parce que ces caractéristiques nous permettent de voir en quoi il peut être rationnel de s'engager, dans la plupart des situations, dans un processus cognitivement plus « léger ».

7.1.2.1 Frugalité cognitive et rapidité procédurale

Les règles simples de décision que sont les heuristiques sont cognitivement frugales si on les compare aux procédures décisionnelles d'optimisation. Par 'frugale', un chercheur

comme Gigerenzer entend le fait de négliger intentionnellement des informations pertinentes (2008 : 20-45). Un agent qui « satisfie », par exemple, n'examinera pas toutes les alternatives disponibles pour en identifier la meilleure, ni même toutes les propriétés pertinentes. Il s'arrêtera à la première option satisfaisante et n'examinera pas les suivantes. D'ailleurs, les attributs recherchés dans une option ne sont pas maximisés ou minimisés. Dès que l'examen d'un attribut révèle qu'il passe le seuil d'acceptabilité, l'agent néglige la distance qui sépare l'attribut de ce seuil. En d'autres termes, ce qui compte est de savoir si l'option obtient la note de passage pour l'ensemble des critères préalablement établis. L'agent peut (et doit) négliger le reste.

Pour ce qui est de la rapidité procédurale, il ne fait pas de doute que l'optimisation prend beaucoup plus de temps, dans la plupart des situations, que l'usage d'une heuristique. Peser le pour et le contre de chacune des options disponibles peut rapidement devenir une quête qui engage toute une vie. Les choix amoureux en sont, je pense, de bons exemples. Rechercher toutes les informations pertinentes pour déterminer laquelle de deux soupirantes doit avoir les faveurs de notre cœur peut durer de nombreuses années. Dans des cas de ce genre, les informations pertinentes ne sont connues que par fréquentations répétées des options dans des situations variées susceptibles d'être étalées sur une longue période.

On peut mentionner également toutes les situations de choix dont les solutions optimales existent, mais sont computationnellement irrésolvables – ou du moins pas sur une période temps raisonnable. Les théoriciens des jeux savent que les échecs, par exemple, comportent une solution. Pour tous les tours, il y a au moins un meilleur coup. Hypothétiquement, un joueur n'aurait qu'à élaborer un arbre de décision comprenant toutes les possibilités de coup et partir des résultats qu'il souhaite obtenir pour remonter la chaine jusqu'à sa position actuelle. Il sélectionnera alors le coup qui comprend le résultat qu'il vise dans son arbre. Or, comme les parties possibles aux échecs sont de l'ordre de 16^{64}, ce genre de démarche s'étalera sur plusieurs millions, voire plusieurs milliards, d'années! De même, rechercher l'itinéraire le plus court entre plusieurs destinations peut rapidement devenir un problème computationnellement insoluble. Comme la quantité de possibilités est exponentielle, on a montré que les plus puissants ordinateurs ne peuvent résoudre dans un temps raisonnable le problème d'identifier le trajet le plus court pour un touriste qui souhaiterait visiter les capitales des 50 états états-uniens.

Ce que ces cas montrent, c'est qu'on peut être certain qu'il existe un meilleur choix dans certaines situations, mais le temps nécessaire pour le découvrir peut nous amener à y renoncer. Vaut mieux parfois nous contenter d'une solution sous-optimale que de nous engager dans une délibération complète pour résoudre un problème décisionnel.

7.1.2.2 Flexibilité dans l'ordre partie/tout

Les procédures classiques de formation du meilleur jugement prescrivent un examen de l'ensemble des alternatives disponibles et de l'évaluation de chacune d'elle à l'aide de critères positifs et négatifs. Ces critères ne sont pas nécessairement connus au départ. Un processus de délibération complet devrait ternir compte également des aspects originaux que comporte une option pour l'évaluer correctement. Aussi, dans la formation d'un meilleur jugement, les agents doivent « partir » des alternatives disponibles et ensuite identifier successivement leurs attributs pertinents. Une fois ces deux étapes réalisées, un agent rationnel devrait effectuer un calcul d'utilité avec la méthode de son choix.

L'ordre des deux premières étapes est important. On parle de priorité tout/partie lorsqu'on identifie l'ensemble des alternatives avant d'identifier leurs attributs. Mais certaines heuristiques décisionnelles renversent cet ordre. Satisfier, par exemple, implique que l'on sache au départ ce que l'on souhaite retrouver dans une option avant de commencer notre recherche d'alternatives. On débute donc par sélectionner des attributs et ensuite, on sélectionne ou élimine des alternatives. Dans ce cas, on parlera d'une priorité des parties sur le tout. Par exemple, rechercher efficacement une propriété immobilière requiert que nous sélectionnions d'abord ce que nous souhaitons y retrouver. On formera un seuil avec les critères dont nous disposons. On pourra, par exemple, stipuler que la propriété doit se trouver à l'intérieur d'une zone de 15 km de notre lieu de travail, coûter moins de 300 000 $, avoir un sous-sol habitable et plus beaucoup de rénovation à effectuer dès l'acquisition, etc. On examinera ensuite des offres sur le marché en fonction de nos critères. On se concentrera d'abord sur les propriétés qui se trouvent dans la zone convoitée, puis on éliminera celles qui coûtent plus de 300 000 $, etc.

Toutes les heuristiques ne procèdent toutefois pas de cette façon. Les méthodes lexicographiques peuvent chercher à trancher à l'aide d'un seul attribut – jugé plus valide ou plus saillant que les autres – entre des alternatives dont l'agent connaît déjà l'étendue.

Mais les heuristiques sont plus flexibles à cet égard que les procédures d'optimisation. La raison en est simple : lorsqu'on priorise les parties sur le tout, cela implique nécessairement une recherche frugale – ce qui est incompatible avec l'optimisation.

7.1.2.3 Règle d'arrêt

Ce qui rend les procédures d'optimisation particulièrement difficiles à suivre est le fait qu'elles ne comportent pas de règles d'arrêts toujours évidentes ou même praticables. Vous pouvez savoir que vous devez arrêter votre recherche d'alternative lorsque vous les avez toutes identifiées, mais il est souvent difficile, voire carrément impossible, de savoir quand vous avez effectivement identifié toutes les alternatives. Le magasinage d'une paire de chaussures en est un exemple évident – à moins bien sûr que vous sachiez que votre région ne comporte que deux ou trois marchands de chaussures. En revanche, si vous « satisfiez », vous savez, à chaque étape, quand vous devez continuer votre recherche ou vous arrêter et sélectionner un choix. D'ailleurs, beaucoup de consommateurs modifient en cours de route leurs procédures de choix et éternisent sans trop s'en rendre compte leur déambulation dans les magasins. Une fois qu'ils ont trouvé un item qui se situe au-dessus de leur seuil de satisfaction, ils ne le choisissent pas tout de suite et se disent qu'il serait préférable d'aller voir dans une autre boutique pour s'assurer qu'il n'y a pas un item plus avantageux. Quiconque fait du magasinage avec son conjoint sait à quel point le passage à une procédure d'optimisation est quelque chose d'exaspérant!

Une règle d'arrêt ne concerne pas exclusivement la recherche d'alternatives. Elle peut tout aussi bien s'appliquer à la recherche d'attributs, de raisons ou même de preuves pour évaluer les alternatives. Si je dois, par exemple, déterminer si je dois m'engager dans une relation amoureuse ou demeurer célibataire, cela requiert que j'examine des raisons. Mais est-ce que je dois examiner *toutes* les raisons pertinentes et en évaluer les poids relatifs? *A priori*, je serais tenté de répondre par l'affirmative. Mais en y regardant de plus près, force est de constater qu'il difficile de savoir où je dois m'arrêter dans ce processus. Je peux, par exemple, estimer qu'être célibataire comporte l'avantage de ne pas avoir l'obligation de visiter et d'entretenir des relations avec une belle-famille; cependant, estimer le poids relatif de cet inconvénient est une tâche non seulement ardue, mais suffisamment complexe pour qu'il soit difficile de dire si l'on a assez d'informations pour le faire correctement. Pour être rigoureux, je devrai élaborer toutes sortes d'hypothèses auxiliaires concernant la

quantité et la durée des rencontres familiales, ainsi que la qualité moyenne ou la quantité de déplaisir et de plaisir que je peux escompter avoir en fréquentant ma belle-famille, ou le potentiel qu'elle représente pour la modification de mon propre caractère, etc. En fait, pour chaque problème qu'on soulève dans la délibération, un sous-problème auxiliaire apparaît dans cette même délibération.

D'ailleurs, les problèmes décisionnels complexes comportent tant d'aspects pertinents qu'il est difficile, voire souvent impossible, de savoir si l'on en a fait le tour. Chercher des raisons n'est pas une activité cognitivement plus simple que de rechercher des alternatives dans un monde qui en offre tant. D'où l'idée que nous devons nous appuyer sur une règle d'arrêt claire et réellement praticable par des agents cognitivement limités.

7.1.2.4 Stratégie non compensatrice

Dans une délibération complète, les agents doivent soupeser les avantages et les inconvénients de chacune des options disponibles. Soupeser consiste non seulement à pondérer les options qui s'offrent à nous, mais à les faire se contrebalancer afin de déterminer la « quantité » de poids supplémentaire que représente un avantage ou un inconvénient par rapport à l'autre. Une règle de compensation peut être aussi simple que l'injonction de Franklin qui consiste à dire qu'on doit éliminer les avantages et les inconvénients relatifs aux alternatives concurrentes qui ont sensiblement *le même poids*. Un calcul plus rigoureux consiste à attribuer une cote négative aux inconvénients et une cote positive aux avantages que présentent les options, laquelle cote est alors multipliée par un facteur probabiliste comme dans les théories modernes du choix rationnel. On additionne toutes ces cotes et l'on sélectionne l'option qui obtient le plus haut score. Quand on additionne[76] de la sorte les valeurs des raisons susceptibles de motiver une décision, on a alors affaire à une stratégie compensatrice.

Or, les heuristiques décisionnelles ne comportent pas en général de règles de compensation de la sorte. Ces règles de compensation sont cognitivement plus exigeantes et participent clairement au caractère peu frugal des procédures d'optimisation.

[76] Tout dépendant de la méthode de calcul utilisée, on peut aussi soustraire.

Une procédure comme « satisfier », par exemple, ne fait pas jouer les attributs les uns contre les autres. Les critères qui entrent dans la confection d'un seuil sont respectés si et seulement si une note de passage est obtenue par une option pour chacun de ces critères. Il n'y a donc pas de calculs linéaires d'addition et de soustraction des valeurs des attributs pertinents.

La même chose vaut pour les méthodes lexicographiques comme la sélection ou l'élimination par attribut. Si je cherche à me procurer une automobile neuve, il sera judicieux que je fixe au départ une série de critères qui me permettront de trancher entre plusieurs véhicules. Je peux stipuler que la voiture doit coûter en dessous de 20 000 $. J'éliminerai alors une quantité très grande de modèles. Mais s'il en reste, alors je devrai utiliser un autre critère dans l'espoir qu'il me permettra de trancher entre les modèles restants. Je peux alors stipuler que la voiture doit avoir reçu une note de fiabilité élevée dans les grandes revues spécialisées. S'il reste deux modèles, je procéderai alors à leur examen à l'aide d'un nouveau critère, par exemple la consommation d'essence ou la beauté relative, etc. Mais même si j'ai utilisé ici plusieurs critères pour solutionner mon problème décisionnel, aucune solution de compromis ne fut nécessaire dans l'évaluation des options : la valeur d'*un seul* attribut *comptait* à chaque étape du processus décisionnel.

Pour chaque question que l'agent se pose à une étape du raisonnement pratique, il doit fournir une réponse à l'aide d'un seul critère *sans égard* aux réponses subséquentes qu'il obtiendra en usant d'un autre critère si sa recherche devait se prolonger. Comme le coût d'acquisition d'informations pertinentes est souvent important, la sélection ou l'élimination par attribut permet à l'agent d'éviter de s'engager dans une recherche trop coûteuse – à condition bien sûr que les attributs utilisés soient les meilleurs ou les plus adaptés.

En revanche, un agent maximisateur n'éliminera aucune alternative en partant. Par exemple, il n'exclura pas d'un examen subséquent les automobiles de plus de 20 000 $ parce qu'il estime avoir des chances de dénicher une voiture un peu plus chère, mais qui offre beaucoup plus de fiabilité, d'efficacité de consommation et de qualités esthétiques. Ce faisant, il devra pondérer la valeur relative de ces attributs et le coût supplémentaire qu'ils représentent. Aussi, un agent qui utilise une méthode lexicographique renonce d'une certaine manière à la meilleure option. Au terme de sa recherche, il peut estimer qu'il n'a

probablement pas identifié *la* meilleure option, mais ne s'est assurément pas engagé dans une recherche trop coûteuse.

7.1.2.5 Combinaisons

Les heuristiques de décision sont variées, mais souvent assez compatibles pour être combinées. Les grands maîtres d'échec, par exemple, n'examinent pas une quantité très grande de stratégies et de coups, mais ils en identifient d'abord trois ou quatre qui leur semblent satisfaisants à l'aide de critères généraux préétablis, puis essayent de trouver le meilleur parmi ceux-ci en imaginant les issues les plus probables pour chacun d'eux, pour finalement choisir celui qui leur apparaît le meilleur (Klein : 2001). Des recherches ont même mis en relief que les grands maîtres optent la plupart du temps pour l'un des premiers coups qui leur viennent à l'esprit lorsqu'ils observent une position. Ceci suggère qu'un ancrage décisionnel joue également un rôle dans la prise de leur décision. De même, les choix des consommateurs sont souvent dirigés dans un premier temps par l'heuristique de la reconnaissance. Ils éliminent spontanément les items dont ils ne connaissent pas la marque, et utilisent ensuite une méthode lexicographique ou la méthode d'élimination des extrêmes pour faire le reste du travail.

En revanche, un agent optimisateur n'a pas cette latitude. Il doit choisir une seule procédure et s'y tenir du début du processus à la fin. Les procédures d'optimisation sont en ce sens « exclusives ». La raison profonde de cette exclusivité réside dans le fait qu'elles ne sont pas frugales, et qu'elles doivent tenir compte de toutes les informations pertinentes.

7.2 Maximiser ou utiliser une heuristique : le choix rationnel

La formation d'un meilleur jugement est classiquement vue comme l'exercice par excellence de la rationalité. Les auteurs classiques et modernes estiment typiquement qu'un choix véritablement rationnel repose nécessairement sur un examen complet des tenants et aboutissants de toutes les alternatives disponibles. Bien qu'il s'agisse d'un réquisit que certains estiment être trop contraignant, il y a tout de même un préjugé favorable, mais erroné, en faveur des procédures d'optimisation. D'ailleurs, il existe depuis les années 1970 tout un programme de recherche, initié par Kahneman et Dverski, orienté dans l'identification et la description d'illusions cognitives produites par l'usage de règles de

décision et de jugement simple, telles que les heuristiques. Ce programme vise à mettre en relief seulement les situations où les agents qui utilisent des heuristiques font des erreurs *systématiques*. Les chercheurs participant à ce programme négligeaient l'étude du caractère utile et adaptatif de ces procédures de décision ainsi que l'étude de la nature des situations dans lesquelles elles s'avèrent être des stratégies gagnantes.

Il a fallu attendre les travaux de Gigerenzer, Todd, Selter, Goldstein et du ABC Group qui s'inspiraient d'une relecture des travaux classiques d'Herbert Simon pour opérer un renversement de paradigme. Ces chercheurs ont démontré sur la base d'expériences et de données expérimentales que l'usage des heuristiques correspond non seulement à la manière dont les agents raisonnent la plupart du temps, mais qu'elles offrent des chances de succès supérieures dans les situations appropriées.

Les résultats expérimentaux sont suffisamment robustes pour récuser quelques idées fausses, mais tenaces, concernant l'usage des heuristiques pour résoudre des problèmes cognitifs et décisionnels. Je pense qu'il est utile de les mentionner ici, si l'on souhaite montrer qu'il peut être rationnel de ne pas s'engager dans la formation d'un meilleur jugement, voire de le négliger même lorsqu'on y a accès.

7.2.1 Quelques idées fausses concernant l'usage des heuristiques décisionnelles

J'emprunte ici à Gigerenzer, *Rationality for Mortals*, la liste des préjugés défavorables à l'égard des heuristiques dans la formation des jugements ou la prise de décision :

(1) Comme les heuristiques produisent toujours des solutions de deuxième ordre, l'optimisation est toujours meilleure.

(2) Notre esprit utilise des heuristiques seulement parce qu'il est cognitivement limité.

(3) Les agents utilisent des heuristiques seulement pour prendre des décisions routières, jamais pour prendre des décisions importantes.

(4) Plus on a d'informations et plus on utilise des ressources computationnelles, plus on a des chances de prendre de bonnes décisions ou de former de bons jugements.

L'énoncé (1) est erroné parce qu'on ne tient pas compte des situations dans lesquelles la formation d'un meilleur jugement n'est pas computationnellement réalisable dans des

délais raisonnables. Mais il est également faux parce que le meilleur jugement manque de robustesse dans la mesure où la pondération des raisons comporte une marge d'erreur trop grande. Cela explique, jusqu'à un certain point, pourquoi nous ne sommes pas convaincus par des arguments sophistiqués et trop longs, *même si nous sommes capables d'en suivre correctement le cours*. L'addition des erreurs d'estimation sape la justesse d'un meilleur jugement.

L'énoncé (2) est erroné parce que nous nous servons des heuristiques pour des raisons qui ont à voir avec la structure du problème, tel que son caractère computationnellement irrésolvable, les écueils qu'il représente pour la robustesse des solutions et son urgence.

L'énoncé (3) est erroné parce que l'importance qu'un problème représente pour nous ne fait pas partie de sa structure. Nous pouvons, par exemple, parier sur le gain d'une équipe de hockey lors d'un match à venir, et dont l'enjeu peut être aussi bien la fortune de toute une vie qu'une petite tape dans le dos venant d'un ami. Les agents rationnels utilisent des heuristiques tant pour faire des choix importants que des choix triviaux.

L'énoncé (4) est erroné parce que, paradoxalement, il peut être judicieux d'ignorer des informations qui nécessiteraient une computation complexe susceptible de saper, encore une fois, la robustesse des résultats. Les comportements d'investissement illustrent bien ce point. L'économiste Harry Markowitz a obtenu le Prix Nobel pour ses travaux sur les allocations optimales des placements entre différents fonds. Il en est venu à la conclusion qu'il est, dans bon nombre de situations, statistiquement avantageux pour un investisseur d'allouer tout simplement ses capitaux également entre ses fonds, *sans égard pour les perspectives différentes qu'offrent ces fonds*[77].

7.2.2 La rationalité écologique

Comme je l'ai mentionné, les recherches actuelles tendent à montrer que la justesse, voire la supériorité, des heuristiques sur les procédures d'optimisation réside en partie dans la structure des problèmes que celles-ci visent à solutionner. Aussi, le fait qu'une heuristique soit adaptée à son problème peut prendre plusieurs formes. Une de ces formes consiste pour un agent qui utilise l'heuristique appropriée à « faire faire » la plus grande partie du travail

[77] Commenté par Gigerenzer (2008 : 8-9).

computationnel par l'environnement lui-même avant de le traiter par ses propres ressources cognitives[78].

Par exemple, Goldstein et Gigerenzer (2002) ont établi que l'heuristique de la reconnaissance offre de bons résultats pour évaluer les populations relatives des grandes villes, les perspectives de gain des équipes sportives et des fonds de placement. Si vous demandez à des citoyens américains de déterminer laquelle de Détroit ou de Milwaukee est la ville la plus peuplée, ils obtiendront des résultats statistiquement moins bons que les citoyens allemands. Contrairement aux Américains, la plupart des Allemands peuvent utiliser l'heuristique de la reconnaissance pour répondre à la question. Ils reconnaissent Détroit, et non Milwaukee, et en infèrent correctement que la première est plus peuplée que la seconde. Pourquoi obtiennent-ils de meilleurs résultats, alors qu'ils en connaissent beaucoup moins sur ces deux villes que l'Américain moyen? La structure de l'environnement est telle que, plus une ville est importante, plus elle fera parlée d'elle dans les médias, et plus on parle d'elle dans les médias, plus il a y aura de chance que des personnes en entendent parler. La distance géographique qui sépare les Allemands du continent américain est telle que si les signaux médiatiques concernant Détroit se rendent dans les foyers allemands, mais pas les signaux concernant Milwaukee, c'est probablement parce que la première est plus importante que la seconde.

De même, satisfier est une procédure particulièrement adaptée aux environnements qui présentent aux agents des options nombreuses et qu'ils ne peuvent qu'examiner de manière sérielle (ex. : les biens de consommation pour lesquels il faut magasiner). Si l'environnement est en plus particulièrement instable, « satisfier » devient une stratégie incontournable (ex. : chercher une maison sur le marché de l'immobilier « usagé »).

[78] Cette hypothèse fut initialement proposée, illustrée et défendue par Herbert Simon au cours des années 1950. Simon croyait que l'efficacité d'une règle de décision et de traitement de l'information devait refléter la structure de l'environnement. Aussi, en examinant cette structure, nous sommes en mesure d'inférer des caractéristiques fondamentales de l'esprit des organismes qui ont du succès dans cet environnement. Par exemple, si l'on observe qu'un organisme évolue dans un environnement qui présente des sources de nourriture distribuées de manière aléatoire, on peut en inférer que l'organisme suit également une règle aléatoire de recherche nutritive. Mais si son environnement comprend des sources constituées en grappe, mais distribuées aléatoirement, ses capacités cognitives seront différentes. Il utilisera initialement une règle aléatoire. S'il tombe sur une source de nourriture, il passera à une règle qui lui commande de rechercher ensuite dans les zones adjacentes à cette source. Cette dernière règle doit spécifier un nombre d'essais avant de retourner en mode aléatoire. La grandeur typique des grappes devrait nous donner des informations sur le nombre d'essais que l'organisme doit effectuer avant de passer en mode aléatoire. Plus les grappes sont petites, plus petit sera le nombre d'essais, etc.

L'imitation est aussi une procédure qui requiert que des conditions environnementales spécifiques soient réalisées pour offrir des chances de succès significatives. L'observateur et le démonstrateur doivent évoluer dans un environnement physique ou social similaire. L'environnement doit être relativement stable. L'imitation est profitable quand l'environnement comporte beaucoup de « bruits » et que les conséquences des lignes de conduite ne sont pas immédiates, donc assez difficiles à évaluer comme les choix politiques et moraux (Goldstein et coll. 2001). L'imitation est d'autant plus appropriée lorsque les coûts de l'innovation sont prohibitifs (Axelrod & Cohen, 1999).

L'efficacité de « *take the best* » ou l'élimination par aspect est bien documentée (Martignon, Hoffrage : 1999). Cette heuristique est « écologiquement » valide quand les critères lexicographiques et leur ordre d'application correspondent à des structures environnementales qui exemplifient des informations non compensables qui se prêtent difficilement à un traitement computationnel linéaire (ex. : sélection de candidat, acquisition d'une maison, choix d'un partenaire sexuel, appel à l'expertise, etc.). La sélection et l'élimination par attribut sont particulièrement adaptées aux environnements qui impliquent pour l'agent des coûts élevés d'acquisition d'information.

Les heuristiques sont aussi utiles dans des environnements qui comportent des aspects non paramétriques, comme les contextes de choix stratégique. La célèbre stratégie *donnant-donnant* décrite par Robert Axelrod illustre bien ce genre d'heuristique utilisée dans des contextes de coopération qui comportent néanmoins des alternatives opportunistes. Elle consiste à coopérer avec autrui lors de l'interaction initiale, puis d'imiter son comportement jusqu'à la fin. Cette heuristique permit entre autres de voir apparaître la coopération dans des milieux aussi hostiles que celui des tranchées de la Première Guerre mondiale. Les soldats de bataillons adverses avaient tout intérêt à coopérer en feignant de tirer pour tuer, mais ne pouvaient s'assurer que la partie adverse fasse sa part dans les moments critiques. *Donnant-donnant* put tout de même proliférer sur le front, et cela, en dépit des risques de se faire repérer par les officiers. Les conditions environnementales requises pour la prolifération et la robustesse de *donnant-donnant* sont multiples. Elles vont de la capacité commune de punir autrui pour l'absence de coopération à la capacité pour les protagonistes de garder en mémoire la nature et l'initiateur du comportement opportuniste, en passant par

le caractère indéterminé de la quantité d'interactions futures avec les mêmes protagonistes, etc. (Axelrod, 1984).

La variété des heuristiques et leur exploitation de la structure des problèmes font de l'ensemble de ces procédures décisionnelles une véritable boîte à outils adaptative. Il est préférable d'avoir beaucoup d'outils faciles d'usage et bien adaptés aux tâches qu'on souhaite réaliser que d'avoir un seul et unique outil complexe et difficile à utiliser et relativement peu adapté. Il y a des situations qui se prêtent bien à la formation d'un meilleur jugement, d'autres non.

En fait, la plupart des situations de choix comportent des aspects qui rendent difficile l'optimisation. On peut pratiquer l'optimisation lorsqu'on fait face à une quantité très restreinte d'options qui comportent, à leur tour, très peu d'aspects pertinents plutôt stables dans le temps, et dont on peut mesurer aisément et sans urgence le poids relatif.

D'ailleurs, l'optimisation est aussi souvent utilisée pour faire des évaluations *post facto*. Ces évaluations sont parfois complexes, mais peuvent être réalisées dans des délais raisonnables parce qu'on connaît alors l'impact ou le poids relatif qu'ont effectivement eu les divers attributs des choix, ce qu'on n'avait pas comme information lors de l'évaluation *ante facto*. Par exemple, je peux refuser de miser sur un athlète participant à une compétition parce que je viens d'apprendre qu'il vient tout juste de perdre ses parents dans un accident tragique. Après coup, il se peut que l'athlète en question ait été « positivement » affecté par cet événement, en ce sens qu'il a senti naître en lui une force nouvelle et l'urgence de vivre à 100 km/h, sentiment qui, du reste, est en partie responsable de sa première place au classement.

Mais les évaluations *post facto* ne sont pas toujours aussi faciles à produire. Dans les sports et les jeux, l'étendue des coups permis est clairement définie. C'est pourquoi l'issue est rarement contestable. Par contre, dans les débats politiques, les choix existentiels ou esthétiques, il est plus difficile d'identifier un « gagnant ». L'étendue des stratégies disponibles n'est pas clairement définie. Aussi, il est difficile de les évaluer, donc de déterminer s'il s'agit d'un choix optimal.

7.2.3 Maximisation sous contrainte : une avenue praticable?

Après la publication des travaux d'Herbert Simon, beaucoup d'économistes ont tenté de modifier leur conception du choix rationnel en y incluant des considérations sur la contrainte temporelle (Stigler, 1961). De même, Rawls, conscient du problème que l'optimisation représente pour une théorie réaliste du choix rationnel, affirmait à propos du choix d'un projet de vie que

> « […] *la délibération rationnelle est, en fait, une activité comme une autre, et la question de savoir jusqu'où s'y engager est elle-même tranchée par une décision rationnelle. La règle formelle à appliquer est que nous devrions délibérer jusqu'au moment où les avantages probables dus à une amélioration de notre projet compensent exactement le temps et l'effort de la réflexion supplémentaire. Dès que nous prenons en considération les coûts de la délibération, il est déraisonnable de se préoccuper de trouver le meilleur projet, celui que nous choisirions si nous avions une information complète. Il est parfaitement rationnel de suivre un projet satisfaisant dès que le souci de la réflexion n'est pas compensé par les avantages que procureront plus de calculs et d'information.* » (1973 : 459)

J'ai longtemps pensé, comme Rawls, que l'inclusion de contraintes temporelles à la formation d'un meilleur jugement réglait une partie du problème que pose le modèle de la maximisation. Mais beaucoup de chercheurs, œuvrant au sein de l'ABC Group, affirment que la maximisation sous contrainte est en fait une procédure cognitivement plus lourde que la maximisation *tout court*, et ne devait pas, par conséquent, être une bouée de sauvetage pour quiconque souhaite s'engager dans la formation d'un meilleur jugement, alors que la structure des problèmes envisagés ne le permet pas *prima facie*. En discutant avec des collègues économistes, nous en venions cependant à la conclusion que l'optimisation sous contrainte apparaissait comme une bonne solution qui préserverait en partie le modèle d'agent rationnel optimisateur. Aussi, je ne comprenais pas pourquoi certains détracteurs de ce modèle estimaient qu'ajouter des contraintes temporelles à la recherche de la meilleure option alourdissait le processus.

Je pense que je comprends maintenant la teneur de cette critique. Un agent qui examine des alternatives ou des raisons ne peut savoir à quel moment le coût d'un prolongement de cette recherche surpasse les bénéfices escomptés qu'au terme de cette recherche. Comment puis-je savoir, par exemple, que je dois arrêter ma recherche de chaussures à la quatrième paire examinée dans l'éventualité où cette paire constitue un choix optimal *étant donnée* la quantité de ressources que je devrai dépenser pour continuer ma recherche? En fait, rien ne me dit que je ne dénicherai pas une paire de chaussures beaucoup moins chère, plus belle,

beaucoup plus durable et confortable dans la prochaine boutique inscrite sur ma liste. Il est toujours possible qu'avec un minimum d'énergie supplémentaire, je tombe sur un choix dont la supériorité surpasse de loin les choix antérieurement examinés. On ne peut donc être capable d'identifier le meilleur choix, compte tenu des ressources qu'on doit y investir, qu'une fois l'ensemble des alternatives et des raisons connues.

Qui plus est, on doit également colliger des informations sur les ressources utilisées à chaque étape de la recherche *et* calculer, une fois la recherche achevée, le moment où *on aurait dû* faire un choix. Cela alourdit considérablement le processus de délibération.

Je pense que ce qui précède constitue une bonne raison pour croire que l'optimisation sous contrainte n'est pas une avenue praticable pour ceux qui estiment qu'un agent rationnel agit nécessairement en conformité avec son meilleur jugement. Le coût de formation d'un meilleur jugement est, dans la plupart des situations, si prohibitif qu'un agent rationnel doit se contenter d'une solution sous-optimale.

7.2.4 La maximisation « *comme si* »

Les partisans de l'optimisation peuvent concéder que leur modèle n'est pas réaliste, mais que les agents se comportent tout de même « *comme si* » ils optimisaient. Donc, on peut continuer à parler d'agents qui forment de meilleurs jugements, et qui, tantôt s'y conforment, tantôt ne s'y conforment pas. La conception Standard de la stricte akrasie peut être de cette manière préservée et garder toute sa portée descriptive.

Il n'est pas dans mon propos ici de réfuter cette affirmation. Je n'ai pas l'espace ni même les compétences requises pour le faire ici. Il s'agit d'une question empirique plus que philosophique, bien que les conséquences philosophiques soient, à mon avis, importantes. Les débats entourant la nécessité et la portée du principe de charité dans la description et l'évaluation des choix et du comportement des agents sont souvent tributaires d'une conception « comme si » de la rationalité. On postule que les agents sont rationnels au sens où ils se comportent *comme s'*ils maximisaient la satisfaction de leurs préférences et minimisaient la réalisation de leurs aversions. Cette conception est aussi à la base de la théorie des préférences révélées, telles que décrites au chapitre 3.

Le problème que les chercheurs du programme sur les heuristiques comme boîte à outils adaptative voient dans cette conception est qu'elle laisse le processus décisionnel, et celui de formation du jugement, complètement opaque. Nous ne savons pas vraiment comment les agents traitent les informations, mais nous connaissons le résultat de ce processus de computation et nous postulons *a posteriori* qu'il est compatible avec un processus d'optimisation, et qu'il est donc approprié (ou même normativement nécessaire) de le considérer comme issu de ce type de processus. Contrairement aux apparences, cette conception de la rationalité « *comme si* » relèverait non pas d'une sorte d'inférence à la meilleure explication – ce qui serait en fait une bonne chose –, mais une manière de couper court à la recherche d'explications détaillées et empiriquement mieux appuyées[79].

En revanche, décrire les comportements et les procédures décisionnelles en mettant en relief l'usage d'heuristiques nous permet de les rendre plus transparents, donc plus faciles à prédire. Si je postule qu'un consommateur, qui magasine pour une paire de chaussures, optimise, il sera, par exemple, difficile pour moi de prédire quand il arrêtera sa recherche et pour quelle paire il optera, *même si je connais l'ensemble de ses critères de choix et l'ensemble des alternatives disponibles pour lui*. Par contre, si je connais le seuil de satisfaction qu'il s'est fixé, je pourrai non seulement prédire quelle paire il choisira s'il l'a rencontré sur son chemin, mais également les différentes paires *qu'il choisirait* s'il empruntait tel ou tel itinéraire.

Considérons un autre exemple, celui d'un choix binaire dans une situation où l'on demande aux agents soit de sélectionner l'une des options ou d'en éliminer une. Du point de vue d'un modèle d'optimisation, le résultat est censé être le même. Conformément à la méthode de Franklin, nos décisions rationnelles devraient dépendre – du moins partiellement – des poids que nous assignons aux avantages et aux inconvénients de celles-ci. Toutefois, comme Shafir, Simonson et Tversky le font observer, les avantages (ou attributs positifs) d'une option pèsent plus lourd dans la balance lorsque nous nous proposons de *sélectionner* un choix, tandis que ce sont les inconvénients (ou attributs négatifs) qui pèsent plus lourds

[79] À cet égard, on peut considérer les modèles de rationalité « *comme si* » comme ce que Bachelard appelait des obstacles épistémologiques, à savoir « des causes de stagnation et même de régression » de l'enquête scientifique. Parce que « la connaissance du réel, nous dit Bachelard, est une lumière qui projette toujours quelque part des ombres ». (1938 : 5). Mais il n'est pas toujours facile de distinguer, dans l'imputation de comportements rationnels, les raccourcis épistémiques utiles des étapes trop vite brûlées.

lorsque nous nous proposons d'éliminer un choix (Shafir, Simonson & Tversky, 1993). Si l'on prend en considération la remarque de ces chercheurs, on peut prédire par exemple le choix d'un jury qui doit trancher pour attribuer la garde d'un enfant à un des deux parents en instance de divorce. S'il y a une option *enrichie* (ici, la garde par le parent A) et une option *appauvrie* (ici la garde par le parent B), donc une asymétrie dans le degré relatif de la saillance des attributs positifs et négatifs, on peut s'attendre à ce que les jurys soient disposés à sélectionner *et* éliminer *le même* choix.

À un échantillon de 170 personnes, on a demandé de se mettre dans la peau d'un membre de jury qui doit trancher sur la base de ces seules informations :

Parent A : revenu moyen, santé moyenne, quantité moyenne d'heures de travail, relation raisonnable avec l'enfant, ayant une vie sociale relativement stable.

Parent B : revenu au-dessus de la moyenne, très bonne relation de proximité avec l'enfant, ayant une vie sociale extrêmement active, voyage beaucoup pour son travail, problèmes de santé mineurs.

L'échantillon était divisé en deux groupes dont le premier avait pour tâche d'attribuer la garde d'un enfant unique, et le second la tâche de dénier ce droit. Du premier groupe, 36 % souhaitaient attribuer la garde au parent A, tandis que 64 % souhaitaient l'attribuer au parent B. Du second groupe, 45 % des personnes déniaient la garde au parent A, mais 55 % la déniaient au parent B. Ici, l'option enrichie est dans une plus grande proportion choisie lorsque l'agent décideur doit *sélectionner,* mais aussi *éliminer*.

Ces résultats sont explicables par l'usage d'une heuristique dont le résultat est prévisible. Mais si l'on persiste à croire que les agents se comportent ici *comme s'*ils étaient des maximisateurs dans la formation de leur jugement, alors on devrait faire toute sorte d'hypothèses plus ou moins gratuites sur leurs préférences de manière à préserver une certaine cohérence interne. Or, même si ces hypothèses s'avéraient être sophistiquées, on serait toujours contraint de les faire *post facto*. Tandis que ces résultats sont prévisibles si l'on adhère à la théorie de l'asymétrie des options enrichies et appauvries.

Je ne peux pas, du reste, passer sous silence la richesse descriptive qu'offre l'hypothèse que les agents rationnels utilisent une boîte à outils adaptative qui comprend plusieurs

heuristiques. Voici un exemple qui, je pense, illustre le mieux cette hypothèse, à savoir celle de l'interception des projectiles en mouvement. Critiquant Richard Dawkins, qui avançait l'idée qu'un homme qui attrape une balle en mouvement « se comporte *comme s'il* avait résolu inconsciemment une série d'équations différentielles pour prédire la trajectoire de la balle », Gigerenzer offre une description des mécanismes réels qui fait l'économie de problèmes computationnels si complexes que même les plus puissants ordinateurs sont actuellement incapables de les résoudre. Pour qu'un joueur de baseball, par exemple, puisse résoudre ce genre de problème computationnel, on doit supposer que

> « […] *le joueur connait intuitivement les familles de paraboles, parce que, en théorie, les balles ont des trajectoires paraboliques. Mais pour sélectionner la bonne parabole, le joueur a besoin d'être équipé d'organes sensoriels qui peuvent mesurer la distance initiale de la balle, sa vélocité initiale, et l'angle de projection. Aussi, dans le monde réel, influencée par la résistance de l'air, du vent, de sa rotation, la balle n'effectue pas une parabole. Par conséquent, le joueur doit, en plus, être capable d'estimer la vitesse et la direction du vent pour chaque point de la trajectoire de la balle, dans le but d'en computer le chemin résultant et le point où la balle tombera, et estimer la course à effectuer jusque-là.* » (1976 : 96)[80]

Pour Gigerenzer, ce genre d'explication repose sur une conception idéale d'omniscience et d'omnipotence des agents. Pour résoudre un problème computationnellement aussi complexe, les agents doivent élaborer sur une période extrêmement brève une représentation complète de leur environnement – c'est-à-dire de connaître la totalité des facteurs pertinents pour résoudre le problème – et être capable de computer efficacement l'ensemble de ces représentations. Or, McLeod et Dienes (1996) ont montré que ce n'est pas du tout ce qui se passe lors que des organismes interceptent des objets en mouvement. Les joueurs de baseball, par exemple, fixent du regard la balle en mouvement dans un angle précis et courent dans sa direction *en maintenant autant que possible l'angle du regard constant*. Les joueurs ajustent donc en cours de route leur vitesse en fonction de cet angle. Les joueurs appliquent dans ces situations l'*heuristique du regard*. C'est une heuristique qui est particulièrement utile pour les organismes qui doivent intercepter leur proie – une variante de cette heuristique, celle qui consiste à se comporter de manière à ce que l'angle du regard se modifie sans que la tête bouge, est cependant utile pour éviter des prédateurs. Ce genre d'explication contraste avec les explications « *comme si* ». Elle est, au point de vue de la description, plus riche parce qu'elle nous donne des informations sur l'ensemble

[80] Cité par Gigerenzer (2008 : 21).

des comportements ou les étapes processus, et non pas seulement sur le résultat final. Un agent qui applique l'heuristique du regard accélèrera sa course à mesure qu'il se rapprochera de la balle, ce qu'on observe, mais qu'on peut maintenant expliquer. Si l'agent évaluait vraiment le point d'impact de la balle, il courrait aussi vite que possible vers ce point, ce qui n'est pas le cas. D'ailleurs, cette heuristique n'explique que la manière dont les joueurs s'y prennent pour attraper une balle qui descend. Ils se servent d'une autre heuristique pour déterminer comment ils doivent se comporter lorsque la balle opère une trajectoire ascendante. Si la vitesse de la balle *semble accélérer* du point de vue du joueur qui doit l'attraper, c'est que celle-ci va atterrir derrière lui. Si elle *semble décélérer*, c'est qu'elle atterrira devant lui. Et si la vitesse *semble constante*, alors elle atterrira dans sa zone. Aussi, le joueur de baseball se sert d'une heuristique différente pour orienter ses décisions à l'égard de deux phases distinctes de la trajectoire d'une balle en mouvement. Or, il appert que les détails qu'on observe dans le comportement des joueurs, s'ils s'expliquent aisément par l'usage de règles de décision simple, restent cependant gommés par un modèle de maximisation « *comme si* ».

7.3 Conséquences pour la conception Standard

Le bref aperçu, que je viens de donner, des diverses procédures qui orientent les agents dans leur décision a pour but d'ébranler quelques idées reçues concernant la rationalité pratique. Il est pertinent de le faire parce que la définition standard de la faiblesse de volonté repose sur l'une de ces idées reçues. Le processus de maximisation, que j'ai décrit au chapitre 5 et qui représente, je pense, assez bien la manière – souvent confuse et approximative – qu'ont les philosophes et les économistes modernes de concevoir la délibération rationnelle et le lieu où réside la source de l'akrasie, est davantage putatif que réel. Les agents rationnels ne s'engagent que *très rarement* dans un processus délibératif qui aboutit dans la formation d'un meilleur jugement. Leurs limites cognitives et la structure des problèmes auxquelles ils font régulièrement face requièrent la plupart du temps l'usage d'heuristiques dont les caractéristiques sont très différentes des méthodes d'optimisation. Or, les épisodes de stricte akrasie sont en revanche *très prévalents* – ce qui a de quoi nous mettre la puce à l'oreille. Mais comment concilier ces deux faits?

Je pense que l'hypothèse à privilégier est tout simplement qu'on peut vivre des épisodes de stricte akrasie *sans* avoir préalablement formé de meilleurs jugements, *ni même essayé de le faire*. Si c'est effectivement le cas, ce que je pense, alors la définition standard devrait être abandonnée. S'il ne s'agissait que de cas limites, la conception Standard pourrait être maintenue, mais étant donné la rareté des meilleurs jugements et la prévalence des cas de faiblesse de volonté, il serait plus approprié d'élaborer une autre conception.

Mes conclusions ici sont toutefois davantage négatives que positives. Une conception qui présuppose que les choix rationnels relèvent d'une forme ou d'une autre de maximisation, comme cela est présupposé dans la conception Standard, manque non seulement de fondements empiriques, mais nous pousse théoriquement à surdiagnostiquer des cas de faiblesse de volonté. Je ne propose pas ici une conception alternative, mais identifie plutôt une limitation sérieuse de la conception Standard et une piste de solution pour les cas de stricte akrasie.

7.3.1 De la stricte akrasie sans meilleur jugement

Un agent peut décider d'adopter un comportement qui relève de la faiblesse de volonté même s'il n'a pas formé de meilleur jugement. Sa décision peut aller à l'encontre d'un jugement dûment formé par l'application d'une heuristique décisionnelle et non pas suite à une pondération complète des raisons qui sous-tendent l'ensemble des alternatives qui s'offrent à lui.

Comment les gens savent-ils, par exemple, qu'il est irrationnel de fumer?

Il y a évidemment beaucoup de considérations qui entrent en jeu lorsque vient le temps de se décider ou de s'abstenir de fumer, et cela, autant pour les personnes qui ont acquis l'habitude de fumer que pour celles qui se demandent si elles devraient commencer à fumer. On devient habituellement fumeur pour toutes sortes de raisons et on demeure fumeur pour d'autres raisons. On commence à fumer pour faciliter son intégration à un groupe social ou s'en distinguer, ou pour faire de nouvelles expériences. On continue à fumer pour des motifs viscéraux comme le manque ou pour retrouver un réconfort dans une source de plaisir assuré et difficilement remplaçable.

Bref, les personnes ont toutes sortes de raisons pour fumer, mais ne considèrent pas habituellement qu'il s'agisse de bonnes raisons. Cela est particulièrement évident pour les fumeurs invétérés qui mettent en garde les plus jeunes du danger de débuter une consommation régulière pour des motifs sociaux. Comme les considérations sociales sont particulièrement importantes pour l'adolescent, les mises en garde sont souvent ignorées. Et puisque les adolescents ont en général une aversion plutôt limitée aux risques, les injonctions relatives à la santé n'ont pas le même impact que chez un adulte.

Sur quoi se basent les personnes informées pour considérer que, malgré tout le plaisir que procure la cigarette pour un fumeur, il est préférable de ne pas fumer? On est tenté de croire qu'ils font reposer leur jugement sur une considération complète des tenants et aboutissants de l'inhalation de fumer et de la métabolisation de la nicotine. Or, il n'en est rien. Les fumeurs ne considèrent que certaines informations et ignorent le reste. Ils considèrent l'avis des experts en santé publique et négligent par exemple de calculer le véritable rapport coût/bénéfice de fumer *à l'occasion* quelques cigarettes. Examinons ici ces deux démarches épistémiques.

Comme on l'a vu brièvement, déterminer si je dois tenir compte d'une information à l'aide d'un appel à l'expertise revient à utiliser une méthode lexicographique. Nous utilisons un seul critère à la fois pour trancher. Pour ce qui est de la cigarette, nous utilisons un seul critère, celui de l'opinion de l'organisation de santé publique du pays. Ce seul signal motive, la plupart du temps, notre adhésion à la croyance qu'il est préférable de s'abstenir de fumer même si nous ne vivons pas actuellement de problèmes de santé significatifs. Dans la formation de ce jugement, nous négligeons évidemment l'examen des raisons qui poussent les autorités de santé publique à faire cette injonction. Nous ne connaissons habituellement même pas le pourcentage de risque que représenterait notre consommation de cigarettes d'avoir des problèmes de santé sérieux, et si cela est compensé par le plaisir certain que comporte nos habitudes de consommation. D'ailleurs, les études scientifiques les plus rigoureuses restent des études génériques. Elles n'offrent qu'une partie de l'information pertinente pour déterminer si je dois continuer à fumer, *étant données les conditions idiosyncrasiques qui sont les miennes* (ex. : le stress relié à mon emploi, la nature de mes loisirs, la quantité de plaisir que je retire de la consommation de cigarette, la quantité de cigarettes que je consomme, etc.). Mon jugement, selon lequel il est préférable

que je ne fume pas, ne tient pas compte de toutes les informations pertinentes, mais souvent d'une seule pièce d'information, l'opinion d'une organisation de santé publique.

L'heuristique utilisée alors par les agents ici est une instance de « *take the best* » :

> *Si les organismes de santé publique, dignes de confiance, font des injonctions, prends des décisions compatibles avec elles.*

Mais fait-on un choix rationnel si l'on ne considère que cette information et qu'on néglige le reste? L'information que nous – en tant que non-spécialistes – considérons exclusivement pour condamner la cigarette est adéquate *parce qu'elle reflète en quelque sorte la structure du problème*.

Je pense qu'on peut décrire les éléments de cette structure de la manière suivante :

(1) Les informations pertinentes ne sont accessibles qu'à ceux qui possèdent déjà des compétences épistémiques poussées et maîtrisables qu'après un long entraînement.

(2) Ces informations sont néanmoins synthétisables et transmissibles par des réseaux sociaux fiables ou dignes de confiance.

(3) La synthèse de ces informations – qui fait fi des anomalies mineures et des idiosyncrasies – augmente paradoxalement la probabilité que celles-ci soient correctes.

Je vais expliquer ici chacun de ces éléments.

L'élément (1) justifie en fait une saine division du travail épistémique dans un monde social où il est difficile d'acquérir certaines informations importantes et où on le fait au prix de recherches méthodologiquement lourdes et d'une formation poussée.

À cet égard, on peut dire que fumer la cigarette est une option particulière sous au moins deux aspects : fumer a une bonne propension « addictive » et comporte des conséquences néfastes, subtiles pour un avenir proche, mais *importantes pour un avenir lointain, et, qui plus est, incertaines*. La connaissance du caractère « addictif » s'établit relativement facilement par « fréquentation ». Tous les fumeurs sont capables de reconnaître les effets du manque de nicotine (ex. : maux de tête, irritabilité, bouche pâteuse, etc.). Par contre, les fumeurs et les personnes en général ne sont pas capables de diagnostiquer un cancer du

poumon, ni d'attribuer une telle maladie à l'usage répété de tabac. L'information pour faire ces diagnostics est en quelque sorte « extractible » de l'environnement, mais ils n'y ont pas directement accès – à moins d'avoir fait de longues études en oncologie.

L'élément (2) consiste dans le fait que même s'il est difficile pour un agent d'acquérir par lui-même l'information pertinente, celle-ci est en revanche disponible par l'intermédiaire des experts. Qui plus est, si le réseau des experts est digne de confiance, l'information qui en provient sera statistiquement plus valide. La confiance est nécessaire parce que les experts ont toujours la possibilité de faire des choix épistémiques peu rigoureux pour toutes sortes de raisons. Et cette confiance repose la plupart du temps sur la réputation de ceux-ci ou des organismes pour lesquels ils travaillent.

L'élément (3) est un peu plus complexe que les deux autres. Un organisme de santé publique fait reposer ses injonctions essentiellement sur deux familles de considérations : des données statistiques et expérimentales, d'une part, et des imputations de préférences (ex. : pour demeurer en santé, pour vivre longtemps, par aversion au risque assez grand, etc.) et autres considérations sur le style de vie moyen des fumeurs (ex. : niveau de stress moyen, niveau de consommation moyen, etc.), d'autre part.

Comme pour tous les problèmes complexes, les données statistiques et expérimentales comportent des anomalies. D'ailleurs, ces anomalies sont souvent mises en relief par des chercheurs qui se présentent souvent, mais pas toujours, comme des rebelles ou des « underdogs » de la science. L'idée de considérer exclusivement l'opinion des autorités publiques en matière de santé est justifiable dans la mesure où ces autorités reflètent un *relatif consensus* parmi les spécialistes, lequel consensus « gomme », pour ainsi dire, les anomalies et l'opinion de chercheurs rebelles plus intéressés par la quête de prestiges sociaux que par le caractère rigoureux de leurs propres recherches. Évidemment, quand les anomalies sont trop importantes pour être négligées, on s'attend à ce que les autorités de santé publique modifient leurs injonctions. Mais, à défaut d'une telle modification, on peut estimer que l'injonction reflète un consensus suffisamment important pour qu'il soit justifié de négliger certaines informations contradictoires.

Pour ce qui est des informations idiosyncrasiques, il est préférable de se fier à l'imputation que les organismes de santé publique font à l'égard de préférences, d'aversions et de styles

de vie, que de faire soi-même ce calcul. Les moyennes établies qui servent de base à ces imputations font fi de différences idiosyncrasiques qui peuvent peser lourd dans la délibération. Mais il peut être préférable de ne pas s'engager dans ce genre de délibération et de considérer plutôt que l'imputation de préférences, d'aversions et de styles de vie par les organismes de santé publique, jusqu'à preuve du contraire, est correcte.

Il est clair que les agents qui forment le jugement qu'ils ne doivent pas fumer sur la base exclusive de l'injonction des organismes de santé publique ne connaissent pas nécessairement les facteurs qui rendent leur règle de décision valide. Ils ne font que l'utiliser. Aussi, quand ils vivent un épisode de faiblesse de volonté, ils prennent une décision non pas à l'encontre de leur meilleur jugement – lequel est d'ailleurs inexistant –, mais à l'encontre d'un jugement qu'ils ont formé en appliquant une heuristique, l'heuristique qui consiste à se conformer aux injonctions des organismes de santé publique.

Il s'agit ici d'un exemple parmi d'autres de cas de formation de jugement formé à l'aide d'une heuristique qui sert de point de référence rationnel pour diagnostiquer un cas de faiblesse de volonté. Le cas présente certes des caractéristiques complexes, mais je pense qu'on peut découvrir des processus cognitifs analogues dans la plupart des cas de faiblesse de volonté. Que ce soit de penser que nous devrions ramasser les feuilles mortes l'automne, changer la pile du détecteur de fumée, s'abstenir de manger des friandises, faire nos travaux scolaires à telle heure, enfourcher notre vélo au lieu de prendre notre voiture ou prendre nos antibiotiques jusqu'au bout de la prescription, rarement nous procédons à une évaluation complète des avantages et des inconvénients de chacune des options disponibles avant de prendre une décision. Nous considérons la plupart du temps certaines informations et négligeons le reste. Je pense que l'exemple de la cigarette est un contre-exemple convaincant.

7.3.2 Des jugements plus adéquats comme point de référence dans la stricte akrasie

Si ce qui précède est correct, alors il devient difficile de maintenir une version de la conception Standard qui fait référence au meilleur jugement de l'agent akratique. Le conflit interne que vit typiquement un agent dans des épisodes de stricte akrasie peut être le fruit de l'application de deux procédures décisionnelles dont aucune ne vise la production d'un meilleur jugement.

Si l'on reprend l'exemple de la cigarette, je peux, après avoir été mis en contact avec un stimulus comme un paquet de cigarettes ou des amis qui sortent fumer à la pause, m'imaginer en train de fumer et mesurer l'impact affectif qui s'ensuit. Si ce dernier a une valeur hédonique très positive, je serai enclin à prendre la décision de succomber une fois de plus à mon vice[81]. Mais je peux être en période de sevrage et me rappeler l'injonction de Santé Canada. Comme j'applique la règle qui veut qu'on doive adopter des comportements compatibles avec les injonctions des organismes de santé publique dignes de confiance, je vivrai un conflit. Or, comme je l'ai montré, l'application de cette dernière règle est loin d'être un processus culminant dans la formation d'un meilleur jugement préférentiel. Le jugement à l'effet que je ne dois pas fumer n'est pas un jugement *toutes-choses-considérées*, donc ne peut pas être un meilleur jugement. Par contre, si je succombe et décide de sortir pour fumer, alors je décide d'agir à l'encontre d'un jugement que d'aucuns pourraient considérer comme étant *plus adéquat* ou *plus juste* que le premier jugement.

7.3.2.1 Double processus, conflit interne et conséquence paradoxale de l'usage de procédures de maximisation non adaptées

Dans le cadre du modèle de l'esprit à double-processus (*dual process*)[82], les heuristiques sont souvent classées par les théoriciens du côté des processus automatiques[83]. Mais rien

[81] Dutton et Aron (1974) ont établi que les agents « construisent » dans certaines circonstances leur volonté en utilisant l'heuristique du *proxy* affectif. L'heuristique du proxy affectif appartient à la famille des méthodes lexicographiques ou méthodes à une seule raison discriminante. L'impact hédonique que procure la simulation interne d'événements virtuellement réalisables est le critère qui permet de trancher entre des alternatives également attrayantes. Dutton et Aron ont concocté une expérience ingénieuse dans laquelle deux groupes d'hommes recevaient un numéro de téléphone provenant d'une jeune femme inconnue qui leur demandait de l'appeler ultérieurement. Au moment de la demande, les hommes du premier groupe vacillaient sur un pont suspendu, alors que ceux du second avaient eu le temps de traverser le pont et avaient donc les pieds sur la terre ferme. La jeune femme reçut une quantité significativement plus importante d'appels venant du premier groupe. L'explication de cet écart réside dans le fait que, si l'on veut la réalisation d'un scénario qui se produira dans un avenir plus ou moins lointain, on se base sur les caractères positifs ou négatifs, ainsi que sur le degré de saillance des sensations internes que produit son imagination en fonction de ce scénario. Or, bien que cela soit une procédure décisionnelle relativement fiable, elle comporte des risques de mésinterprétation. Dans l'expérience de Dutton et Aron, les hommes attribuaient à tort la sensation de chatouillement qu'ils éprouvaient à l'interaction qu'ils avaient avec la jeune femme et non à l'interaction qu'ils avaient avec le pont chancelant.

[82] Voir section 5.1.1.

[83] Dans quelle mesure nous pouvons avoir un contrôle conscient sur nos activités pilotées par des heuristiques dépend de plusieurs facteurs, comme la nature de la tâche qu'elles nous permettent d'exécuter, son domaine d'application et son degré de nécessité adaptative pour les organismes qui les utilisent. Par exemple, les heuristiques perceptuelles qui nous permettent de « colmater » les points aveugles dans notre champ visuel au moyen de la reproduction des portions de stimuli qui entourent ces points, sont fixes et inaltérables : que nous le voulions ou non, nos points aveugles sont colmatés de cette façon. Les heuristiques décisionnelles semblent être en revanche plus malléables. Nous pouvons apprendre consciemment comment nous y prendre pour choisir efficacement une propriété immobilière, ou même un conjoint, par l'application d'heuristiques, ou modifier celles qui guident déjà nos décisions dans ces domaines. Mais rien n'exclut, en principe, que certaines heuristiques décisionnelles soient aussi fixes et peu malléables que les heuristiques perceptuelles.

n'empêche leur utilisation en parallèle avec une procédure contrôlée de maximisation. Un agent vivra une sorte de conflit interne s'il utilise en parallèle une heuristique décisionnelle et une procédure de maximisation – par exemple la méthode de Franklin – et les résultats des deux sont incompatibles. Dans des cas de ce genre, le choix rationnel n'est pas toujours celui qui est dicté par la maximisation. Si l'agent utilise une procédure décisionnelle plus simple, mais plus adaptée que des procédures de maximisation, il aura plus chance d'améliorer ses conditions d'existences.

Comme je l'ai mentionné dans la section précédente, il existe un préjugé voulant que plus on considère d'informations plus on maximise ses chances de prendre de bonnes décisions. Ce préjugé est faux parce que les procédures de décision ne produisent statistiquement de bonnes décisions *que* si elles sont adaptées à la structure du problème. Or, il peut arriver qu'un agent forme un jugement à l'aide d'une heuristique à l'effet qu'il doit faire *x*, mais que, en considérant plus d'informations, il arrive à la conclusion qu'il doit faire *y* quand *x* et *y* sont incompatibles. Si l'heuristique est adaptée au problème, alors que ce n'est pas le cas pour la maximisation, il s'ensuit qu'il devrait suivre son premier jugement.

Si ce qui précède est, dans l'ensemble, correct, un agent pourrait paradoxalement s'engager dans un processus de maximisation plus lourd dont le résultat lui prescrirait d'adopter un comportement qui, intuitivement, *relèverait d'un épisode de faiblesse de volonté*. Comment cela est-il possible?

Un agent peut estimer que son dépendance pour les jeux de hasard est mauvaise pour lui parce qu'elle ne lui permet pas de remplir ses obligations familiales. Il peut établir son jugement sur la base de cette seule raison et négliger les autres. Or, il peut, après une longue période d'abstinence, décider d'ouvrir de nouveau le « dossier » et considérer qu'en ne succombant qu'une seule fois à la tentation dans le courant du mois, il maximiserait son bien-être même si cela comporterait des inconvénients pour sa famille. En effet, succomber à une dépendance après une période d'abstinence crée un bien-être plus intense que dans les moments où les épisodes de consommation sont plus rapprochés dans le temps [84].

[84] L'impact hédonique de la consommation d'un bien d'un même type ne décrit pas une courbe croissante en fonction de la durée de la période d'abstinence. Les végétariens disent que leur envie de manger de la viande croît jusqu'à un certain point après avoir pris la décision de ne plus en manger, mais décroît rapidement une fois ce point atteint, jusqu'à passer le seuil de déplaisir.

Évidemment, si je souhaite former un meilleur jugement à l'égard de la possibilité de m'adonner à mon vice, je devrai considérer beaucoup d'autres raisons. Mais il n'est pas clair que, étant donné le fort impact hédonique qu'une consommation intermittente représente, j'arriverai à la conclusion qu'il vaut mieux m'abstenir de faire une excursion dans un casino.

Quand on entame un processus complet de délibération, c'est qu'on ne connaît pas déjà la solution, sans quoi on orientera de manière biaisée la recherche d'information et la pondération des raisons[85]. Il est évidemment mieux de respecter une règle d'abstinence complète qu'aucune règle à l'égard d'un problème d'addiction. Seulement, rien ne s'oppose, par principe, à l'idée qu'une consommation optimale pour un agent aux prises avec une dépendance consisterait en une consommation intermittente, *en dépit de certains coûts importants*. Peut-être qu'une consommation intermittente n'est pas ici une solution optimale. En fait, elle représente un risque important de rechute. Mais rien ne nous permet d'exclure *a priori* l'hypothèse qu'il puisse s'agir d'une ligne de conduite optimale. Or, il est difficile de mesurer le risque réel de rechute pour un individu particulier évoluant dans une situation particulière. En raison d'un manque de robustesse de l'optimisation dans un contexte aussi complexe, la marge d'erreur est importante et surtout lourde de conséquences[86]. Aussi, nous avons tous l'impression qu'un agent aux prises avec une dépendance qui entame un processus de délibération qui comporte entre autres une mesure de ce risque, en plus d'une mesure du plaisir découlant d'une consommation *post*-abstinence, ne se comporte pas de manière rationnelle. Il devrait plutôt se contenter de suivre la règle simple : « *si tu souhaites remplir tes obligations familiales, ne t'engage jamais dans des comportements addictifs* ». Suivre ce genre règle ne permet peut-être pas à un agent aux prises avec une dépendance d'obtenir un niveau de bien-être optimal, mais

[85] Un peu comme une jeune femme qui entamerait une délibération pour choisir entre deux prétendants, mais qui, voyant le résultat qui se dessine, tente de trouver des raisons supplémentaires pour justifier un choix en particulier. En fait, son « cœur » a déjà choisi (Gigerenzer, 2007).

[86] Un joueur compulsif devrait déterminer, entre autres, avant de faire un choix de consommation ou d'abstinence, la longueur des périodes d'abstinence qui offrent un potentiel maximal de bien-être au point de consommation, les contextes familiaux moins vulnérables à une consommation ponctuelle, l'état d'esprit dans lequel il doit être pour être en mesure de « gérer » les conflits familiaux qui en découleraient hypothétiquement, ainsi que celui qui représente le moins de risque pour une éventuelle rechute (ex. : ne pas aller au casino après une dispute ou pendant qu'on vit un sentiment de « spleen », etc.). Or, le manque de robustesse de toutes ces informations et, surtout, les inconvénients que comporte une période de rodage pour en acquérir certaines, et ensuite les ajuster, militent manifestement en faveur de l'usage d'une heuristique qui comporte une clause d'abstinence complète.

elle lui permet certainement d'avoir un niveau de bien-être « satisfaisant » dans l'éventualité où ce bien-être est en partie tributaire du respect de valeurs familiales. L'usage d'une heuristique d'abstinence intégrale requiert de l'agent *qu'il néglige l'examen d'une grande quantité d'informations pourtant pertinentes*. D'aucuns diront que ce genre de règle simple a fait ses preuves. Pourtant, si un agent aux prises avec une dépendance entame un processus de délibération « complet » et arrive à la conclusion qu'il devrait succomber par intermittence à son vice, et s'il prend effectivement une décision dans ce sens, je pense qu'on peut estimer qu'il vit un épisode de faiblesse de volonté, même s'il se conforme à son meilleur jugement.

Le problème ici est de savoir si l'on doit attribuer à l'agent qui succombe à la tentation un jugement plus adéquat – qu'il aura préalablement formé par l'application d'une heuristique adaptée à la situation – ou non. Les agents akratiques vivent souvent des conflits internes. Dans ces circonstances, un jugement plus adéquat lui indique souvent une meilleure direction, mais reste incompatible avec un autre jugement qui lui indique une option plus attrayante. Mais quand un agent qui succombe à la tentation estime que son jugement le plus adéquat *est justement celui qui lui indique cette avenue* – ce qui arrive quand on observe chez l'agent une modification en profondeur de son état d'esprit – sur quoi peut-on se baser pour diagnostiquer un épisode de faiblesse de volonté?

Je pense que peu importe la réponse qu'on doit donner ici – et il y a sans doute plusieurs solutions – elle doit inclure des considérations sur un jugement de référence, qui n'est pas un meilleur jugement, mais un jugement adéquat que l'agent akratique *aurait dû* former *et* auquel il aurait dû se conformer. Cela a évidemment pour effet d'introduire dans notre diagnostic des considérations normatives et pas seulement descriptives – comme dans la définition Standard. Je vais justement examiner dans le prochain chapitre l'hypothèse de Richard Holton qui veut que des considérations normatives soient nécessaires pour tout bon diagnostic de faiblesse de volonté.

7.3.3 Heuristiques et construction des préférences

Les chapitres 4 et 5 portaient sur la nature et la formation des préférences, et les paramètres classiques à considérer pour faire des diagnostics de faiblesse de volonté. Or, quel est le rapport entre la formation des jugements préférentiels et l'usage d'heuristiques? En fait, la

question est celle de savoir si l'on doit toujours considérer les jugements qui ne sont pas issus d'une délibération, mais de l'application d'une heuristique comme d'authentiques jugements préférentiels.

La question se pose parce que, lorsque nous utilisons des heuristiques pour inférer des jugements portant sur ce que l'on doit faire, il est contre-intuitif de dire qu'il s'agit de préférences. Si je magasine une paire de chaussures et que j'utilise un seuil de satisfaction pour piloter ma recherche, je ne pourrai pas dire, sans tordre un peu l'usage courant du terme, que je préfèrerai la première paire que j'examinerai et qui se révèlera être au-dessus de mon seuil. De même, on ne peut pas vraiment dire d'un agent qui imite ses congénères en adoptant une certaine ligne de conduite, qu'il préfère cette ligne de conduite. Pour établir une préférence, l'agent doit comparer des options[87]. C'est en comparant qu'il arrive à établir un ordonnancement plus ou moins complet de ces options en fonction de leur valeur hédonique actuelle, anticipée ou passée. Or, l'usage d'heuristiques implique que l'agent fasse fi de la comparaison ou du moins n'en fasse qu'un usage parcimonieux.

Mais bien que les jugements formés par l'application d'une heuristique ne soient pas à proprement parler des préférences, les agents ont bien souvent besoin de connaître certaines de leurs préférences, même de manière imparfaite, pour être en mesure d'appliquer correctement bon nombre d'heuristiques. Un agent qui utilise une heuristique pour déterminer ce qu'il doit faire pour solutionner un problème décisionnel entame la plupart du temps une recherche avec un stock de préférences de base. « Satisfier » requiert que nous sachions minimalement ce qui nous voulons obtenir ou éviter d'obtenir. Nous élaborons un seuil à partir de nos préférences et de nos aversions et entamons ensuite notre recherche. De même, l'heuristique de l'élimination des extrêmes repose minimalement sur une connaissance des aspects pour lesquels on a une aversion. En examinant le menu d'un restaurant haut de gamme, j'éliminerai les plats trop dispendieux et les plats qui semblent offrir trop peu parce que je n'aime pas payer trop cher et être insatisfait par un plat.

[87] Évidemment, comme tous les organismes complexes, les personnes naissent avec des préférences « par défaut ». Ces préférences ou ces aversions sont des paramètres plus ou moins fixes, comme le sont la peur des hauteurs, l'aversion à la douleur ou le goût pour les aliments sucrés chez les mammifères. Avec un entraînement ou un contact répété avec certains stimuli, on peut arriver à en altérer l'impact ou à les exacerber.

Si l'on s'en tient à une définition stricte des préférences comme un ordonnancement complet des options disponibles pour un agent, alors il est clair que les jugements obtenus par l'application d'une heuristique n'en sont pas. Mais si ce qu'on entend par préférence est simplement ce que nous voulons faire, obtenir ou éviter, alors je ne vois pas d'inconvénient à les traiter comme des préférences. Je pense que c'est souvent en ce sens qu'on dit préférer acheter telle marque de soupe en boîte ou vouloir tel téléviseur *après* avoir utilisé une heuristique décisionnelle. Évidemment, ce genre de préférence s'inscrit mieux dans un *modèle basé-sur-les-raisons* que dans un modèle *basé-sur-les-valeurs*. La manière dont les agents ordinaires décrivent leur raisonnement pratique fait plus de place aux raisons qu'à des valeurs numériques, ou à des poids délibératifs sur lesquels ils opèrent des calculs.

Un chercheur comme Gigerenzer, qui est sans aucun doute le plus grand promoteur du programme de description des heuristiques et de leurs conditions écologiques d'adéquation, avance que les explications par les préférences sont complètement différentes des explications par les heuristiques (Gigerenzer, 2007). Je ne contredirai pas ici son opinion, mais je pense que ce qu'on doit garder à l'esprit est que le terme « préférence » dans le langage courant a un sens non technique qui le rapproche plus du vouloir que de l'ordonnancement d'options.

Ce point n'est sans doute pas très important. Mais il y a une ambiguïté qu'il est utile de mettre en relief, car certains auteurs comme Joseph Heath défendent l'idée que, la plupart du temps, les agents qui vivent des épisodes de faiblesse de volonté ne s'engagent pas dans des lignes de conduite contre-préférentielles – en excluant les cas où les agents sont pilotés par leur système automatique. Si cela veut dire que les agents décident de faire ce qu'ils veulent faire en se guidant par un jugement *quelconque*, alors cela reste compatible avec les procédures décisionnelles que j'ai décrites. Mais si cela veut dire que – en excluant les cas limites – les agents décident de faire ce qu'ils veulent faire en se guidant par un jugement *toutes-choses-considérées*, alors son analyse reste incompatible avec ce que j'ai mentionné dans ce chapitre.

7.4 Conclusion

L'hypothèse de la maximisation, qui remonte en fait jusqu'à Platon, est sans doute le principal élément théorique qui fait apparaître la stricte akrasie comme quelque chose de

paradoxale. Comment les agents pourraient-ils décider de manière contre-préférentielle s'ils tentent toujours de maximiser la satisfaction de leurs préférences? En fait, l'hypothèse de la maximisation n'est plausible que si l'on dispose conceptuellement des seuls désirs et croyances pour expliquer les décisions. Or, lorsqu'on décrit dans le détail les décisions, les séries de décisions et les comportements des agents dans des contextes de choix, on s'aperçoit qu'ils utilisent toute sorte de règles, d'heuristiques et de procédures décisionnelles qui font fi de beaucoup d'informations pertinentes, mais sont suffisamment adaptées aux problèmes pour assurer le succès de l'agent, et d'une meilleure façon que s'il avait maximisé. Il est donc difficile de défendre l'idée que les agents ne peuvent pas décider de manière contre-préférentielle (ou strictement akratique) parce qu'ils sont réputés toujours maximiser.

Réviser l'hypothèse de la maximisation en incluant dans le calcul les ressources temporelles limitées (maximisation sous contrainte) ne permet pas de sauver l'hypothèse, parce que cela implique non pas un allègement, mais un alourdissement de la tâche cognitive des agents. La maximisation « *comme si* » n'est pas non plus une avenue satisfaisante parce sa puissance descriptive, prédictive et explicative est à peu près nulle. Or, l'identification des règles, des heuristiques et des procédures décisionnelles offre une solution de rechange à la maximisation.

La conclusion qu'on doit en tirer est cependant plus négative que positive. Le sceptique est peut-être dans le vrai, il n'y a peut-être finalement pas de décisions contre-préférentielles. Le problème est qu'il ne peut pas inférer cela de l'hypothèse de maximisation, même dans ses versions plus faibles (maximisation sous contrainte et « *comme si* »). De plus, cela a certaines répercussions sur la conception Standard qu'il vaut la peine de mentionner.

La conception Standard fait référence à une estimation (jugement ou préférence) *toutes-choses-considérées*. Or, comme les agents rationnels utilisent la plupart du temps des règles frugales pour décider, ils sont très loin de considérer toutes les informations pertinentes. Ils ne tiennent compte que de quelques informations – et dans certaines situations, une seule – pour trancher entre des alternatives et n'utilisent la délibération que dans de très rares contextes. Aussi, si l'on accepte l'existence de la stricte akrasie, la conception Standard nous amènerait à en sous-diagnostiquer les cas. Il est plus approprié de ne pas faire référence à un jugement *toutes-choses-considérées* ni même à un meilleur jugement dans la

conception Standard parce que, de toute façon, les agents rationnels ne tentent habituellement même pas d'en former.

Les agents utilisent, pour résoudre la plupart des problèmes pratiques auxquels ils font face, des règles, des heuristiques et des procédures décisionnelles *hétérogènes*, dont les résultats peuvent à l'occasion entrer en conflit. Ce genre de conflit est énigmatique dans la mesure où les agents sont réputés décider conformément au résultat d'un seul et unique type de processus computationnel, à savoir la délibération rationnelle au sens classique du terme. Mais lorsqu'on arrive à identifier les procédures en cause, le phénomène devient plus théoriquement compréhensible.

Je délaisse pour l'instant les disputes sur la stricte akrasie pour aborder la conception de Richard Holton. Holton s'est moins intéressé à la stricte akrasie qu'à l'akrasie tout court. Il a formulé une conception qui jouit actuellement d'une assez grande popularité en raison de sa simplicité et de son élégance. Ma critique sera encore une fois plus négative que positive. Je mettrai en relief des difficultés et des limitations importantes de sa conception.

8

Actions planifiées et bris de résolution : la clef pour comprendre la faiblesse de volonté diachronique?

L'irrésolution d'une part, puis l'inconstance et l'instabilité sont le plus commun et apparent vice de la nature humaine

–Charron

8.1 La théorie de Richard Holton

Dans une série d'articles influents, le philosophe Richard Holton a également critiqué l'idée que la faiblesse de volonté relève de décisions qu'un agent prend à l'encontre de son meilleur jugement. Holton ne nie pas l'existence de cas putatifs de stricte akrasie. Ces cas existent, même s'ils sont loin d'être aussi répandus que les philosophes le pensent en général. Les mécanismes de réduction de la dissonance cognitive limiteraient beaucoup la prévalence de ces cas. Aussi, Holton distingue les cas de faiblesse de volonté des cas de stricte akrasie, et soutient qu'une bonne caractérisation de ces derniers ne s'applique pas aux premiers. Holton mentionne une panoplie de cas qui vont à l'encontre de la définition Standard mais qui demeurent néanmoins des cas très intuitifs de faiblesse de volonté. Mais comme il le dit, il ne vise pas à offrir une conception qui remplacerait complètement la conception classique, mais à offrir plutôt une conception complémentaire qui ne s'appliquerait pas aux cas de stricte akrasie.

Pour Holton, une meilleure caractérisation des épisodes de faiblesse de volonté consiste à dire que les agents qui vivent ces épisodes échouent à se conformer à leur intention. Aussi, la faiblesse de volonté apparaitrait « quand les agents sont trop facilement disposés à reconsidérer leurs intentions » (1999 : 1).

Une telle caractérisation du phénomène comporte pour Holton beaucoup d'avantages qui vont au-delà du fait qu'elle corresponde mieux aux cas intuitifs de faiblesse de volonté. Concevoir la faiblesse de volonté comme une sorte de violation d'intention nous permettrait de rendre compte des cas où les agents prennent des décisions dans un contexte de choix

qui présentent des options incommensurables ou pour lesquelles l'agent a une indifférence marquée. La conception d'Holton nous permettrait également d'expliquer autant les cas de faiblesse que les cas de force de volonté, les cas qui ne présentent pas de conflits internes apparents, et même des cas où les agents vivent des épisodes de stricte akrasie sans vivre des épisodes de faiblesse de volonté et des épisodes de faiblesse de volonté sans stricte akrasie. D'ailleurs, et c'est sans doute ici un des points les plus importants, Holton soutient que sa caractérisation lui permet d'expliquer de manière plus adéquate pourquoi on estime communément que lorsqu'un agent vit un épisode de faiblesse de volonté, il se comporte de façon irrationnelle.

Avant de présenter chacune de ces raisons, cependant, je vais examiner la notion centrale que Holton met au centre de son analyse, le concept d'intention. Je ferai un examen critique des arguments de Holton dans la dernière section.

8.1.1 Intentions, plans et résolutions

La conception que défend Holton est largement tributaire de l'analyse que fait Michael Bratman de la notion d'intention. Holton reprend à son compte de larges parts de l'analyse que Bratman a exposée dans son ouvrage désormais classique, *Intention, Plan and Practical Reason*. Je vais également présenter quelques-unes des propositions centrales de Bratman pertinentes pour comprendre la position de Holton.

Pour Bratman, nous sommes des agents planificateurs. Nous avons la capacité de nous donner des objectifs ou de valoriser des états de choses que nous ne pouvons pas remplir ou obtenir directement. Mais nous avons également la capacité d'élaborer mentalement des lignes de conduite plus ou moins complexes, souvent structurées par étapes, pour les remplir ou les faire advenir. Il s'agit de la confection de plans et la formation d'intentions orientées vers le futur. Aussi, des agents planificateurs comme nous possèdent – à divers degré – la capacité non seulement de former des buts difficiles à atteindre directement et de former des plans ou des intentions pour y arriver, mais de se conformer sur une période plus ou moins longue à ces plans et à ces intentions.

Bratman explique que sans les plans et les intentions nous aurions beaucoup de difficultés à nous coordonner dans nos activités avec les autres, et beaucoup de difficulté à nous coordonner nous-mêmes dans nos propres activités. Ce serait le cas si, par exemple, mon

ami et moi souhaitions faire une activité ensemble, mais que cela requérait que nous nous rencontrions préalablement dans un lieu situé quelque part entre chez lui et chez moi. Si je suis incapable de former l'intention de me rendre à un point de rencontre dans le futur et de me conformer à cette intention, nous échouerons à nous coordonner efficacement. Bratman soutient que les intentions sont également nécessaires pour nous coordonner efficacement avec nous-mêmes, et en particulier lorsque nous devons accomplir une série d'activités dans un ordre déterminé. Un plan pour ma journée, par exemple, me donnera des indications sur ce que je dois faire du temps qui m'est imparti pour cette période et dans quel ordre je dois le faire. Les résolutions, en revanche, ne comportent pas toujours des indications sur l'ordre des activités à faire. Dans leur forme la plus simple, elles stipuleront seulement qu'une activité doit faire partie de ma liste des choses à faire, sans spécifier à quel moment je dois la faire (ex. : je prends la résolution de manger plus de légumes chaque jour ou faire de l'exercice quand je peux). Elles peuvent stipuler également qu'un certain type d'activité doit être exclu de ma liste des choses à faire (ex. : manger des sucreries ou visionner des films tard le soir en semaine).

La différence entre les intentions orientées vers le futur et les plans réside seulement dans le degré de complexité de chacun. Les intentions identifient habituellement une activité seulement qu'on s'engage à accomplir dans le futur. Les plans identifient habituellement une série d'activités et, suivant la nature du problème qu'ils visent à solutionner, un ordre de réalisation ainsi que le moment où celles-ci doivent être réalisées. Aussi, les plans correspondent en quelque sorte à une liste (mentale ou inscrite sur un support externe) d'activités dont chaque option est marquée par un ordre de priorité.

L'utilité de former des plans et des intentions ne se résume pas à la résolution de problèmes de coordination intersubjective et intrasubjective. Les résolutions, que Bratman identifie aux politiques individuelles (*intention policy*), visent l'inclusion ou l'exclusion de certaines activités lors de périodes spécifiques et plus ou moins longues de notre vie. Par exemple, si je prends la résolution de faire de l'exercice durant l'année, je m'engage donc à inclure l'activité de faire de l'exercice quelques fois dans la semaine pendant toute l'année – si c'est ce que j'avais bien sûr en tête lorsque j'ai formé pour moi-même cette résolution. Si je prends la résolution de ne jamais manger une deuxième portion de dessert, j'exclus de

l'ensemble des activités que je ferai à l'avenir la consommation d'un dessert dans les périodes où j'en ai déjà mangé un.

Les politiques individuelles que sont les résolutions sont nécessaires parce qu'elles visent non pas des activités isolées, mais des habitudes. On peut décider de s'abstenir de fumer dans la soirée, mais on peut également décider de ne plus fumer *tout court*, auquel cas l'injonction s'appliquera à un type d'activité non indexé au temps. Je pense que la mention de préférences atemporelles – dont l'usage est critiqué par Heath – vise la plupart du temps à identifier un motif pour adopter une politique individuelle prenant la forme d'une résolution : « J'arrête de fumer parce qu'il est préférable de ne pas fumer ».

8.1.1.1 La nature des intentions selon Bratman

L'une des propositions centrales que Bratman défend – dans son ouvrage classique ainsi que dans les ouvrages et articles subséquents – consiste à dire que les intentions, les plans et les résolutions sont des états mentaux qu'un agent peut former sur la base de ses croyances et de ses désirs, mais ne se réduisent pas pour autant à ses attitudes propositionnelles. Les intentions relèveraient plutôt des décisions, et former une intention serait typiquement prendre une décision à l'égard du futur. Beaucoup de théoriciens ont critiqué la nécessité de postuler des états mentaux hétérogènes comme les intentions, alors que les désirs et les croyances font correctement l'ensemble du travail. Bratman soutient que les intentions ne sont pas des désirs et des croyances pour plusieurs raisons[88], mais il y a deux raisons qui sont particulièrement intéressantes pour la théorie de Holton :

1/ Contrairement aux désirs et aux croyances, les intentions demeurent relativement *stables* dans le temps. Solutionner un problème décisionnel requiert la computation d'informations qui demandent du temps. Les informations seront peut-être plus disponibles dans l'avenir, ou même plus traitables d'une façon aussi efficace. Notre état d'esprit peut changer, ce qui nous amènera à considérer certaines informations, par exemple les facteurs de risques que représente une activité, d'une manière significativement différente. Par exemple, une intoxication ou une excitation future induira chez nous une modification profonde de notre évaluation du risque.

[88] Bratman a identifié et discuté ces raisons dans bon nombre d'ouvrages et d'articles. Je ne mentionne ici que les raisons qui sont pertinentes pour comprendre la conception de Holton.

Lorsque notre motivation repose entièrement sur des jugements préférentiels et non sur des décisions antérieures, elle présente une instabilité dynamique importante. Or, une intention dirigée vers l'avenir nous offre une forme de stabilité nécessaire pour atteindre des objectifs qui autrement seraient trop difficiles à atteindre. Ne pas diriger sa vie au moyen des intentions revient à faire du pilotage à vue. Certaines personnes montrent beaucoup de répugnance à prendre des décisions pour l'avenir. Elles font les choses « quand elles le sentent ». Or, leur horizon d'espérance est plutôt limité parce que rien ne leur permettrait de maintenir le cap si ce ne sont des contraintes extérieures imposées par leurs proches. Beaucoup de facteurs expliquent cette répugnance. Certains éprouvent des difficultés à activer des mécanismes de contrôle de soi situés dans leur néo-cortex, d'autres choisissent un mode de vie qui fait une très grande place à la spontanéité, ou d'autres encore ne veulent tout simplement par être le genre de personnes qui se donnent des contraintes pour l'avenir.

Bratman n'estime pas que, lorsque nous le pouvons, nous devons décider pour l'avenir, mais seulement qu'il est parfois approprié de le faire si cela augmente nos chances de succès. Souvent, il est rationnel de ne pas s'engager dans une évaluation des options disponibles dans le futur parce que notre computation requiert des informations qui ne seront disponibles qu'au moment où nous devrons agir. Il sera alors plus approprié de former une intention orientée vers le présent qu'une intention orientée vers le futur. Mais il s'agit ici d'une proposition sommaire. Il peut être dans certains cas rationnel de s'engager dans un processus de formation d'une intention dirigée vers l'avenir, même si l'on sait pertinemment que des informations cruciales ne seront disponibles qu'au moment d'agir, parce que cela a un effet apaisant sur notre excitation actuelle ou que cela nous évitera des regrets futurs en cas d'échec[89].

[89] Une collègue économiste me disait un jour que les institutions bancaires prennent des décisions financières importantes sur la base de prévision du taux préférentiel anticipé des réserves centrales et que les économistes spécialisés dans la prédiction de ces taux se trompent une fois sur deux. Or, ce fait est bien connu des institutions, et pourtant elles continuent de procéder comme si ces prévisions avaient un taux de succès significatif. Je ne vois que deux hypothèses pour expliquer cela : soit les conseils d'administration tentent d'éviter des blâmes futurs de leurs actionnaires en cas de pertes importantes ou ces conseils considèrent qu'il vaut mieux avoir un mauvais plan pour l'avenir que de ne pas avoir de plan du tout. Cette dernière proposition est appropriée dans un contexte stratégique où l'initiative constitue un avantage crucial. Au jeu d'échec, les débutants apprennent rapidement qu'il vaut mieux pour eux de former un mauvais plan que de gérer les problèmes quand ils se présentent. Former un mauvais plan comporte tout de même l'avantage d'acquérir l'initiative dans certaines positions obtenues à la faveur d'une erreur commise par l'adversaire (Euwe, 1969). Je ne sais pas toutefois dans quelle mesure cela s'applique au monde financier.

2/ La formation d'une intention donne lieu à l'activation d'une série de mécanismes de pilotage des actions, mais aussi à une « gestion computationnelle » des moments qui précèdent l'accomplissement de l'action. Pour Bratman,

> « […] *les plans et les intentions tournées vers l'avenir de l'agent, tant qu'il ne les remet pas en cause en révisant son jugement, contraignent et délimitent l'éventail des options qu'il peut rationnellement prendre en compte dans ses délibérations ultérieures et sa planification à venir.* » (1997 : 77)

C'est la dimension de *contrôle* des intentions que Bratman met ici en relief et qui les distingue des préférences. Les intentions (plans et résolutions) servent de filtre de possibilité dans un traitement computationnel[90]. Par exemple, je peux planifier de travailler toute la soirée au bureau sur un chapitre d'un manuel scolaire que je rédige. La décision de travailler toute la soirée filtre, en partant, toute sorte de possibilités, comme naviguer sur Internet pour le plaisir, discuter, avec des collègues, de problèmes philosophiques qui n'ont rien à voir avec les problèmes que j'aborde dans mon manuel, décider d'aller prendre une bière, visiter régulièrement la machine distributrice de friandises, etc.

De plus, les intentions filtrent les stimuli qui peuvent jouer le rôle de sources de distraction ou filtrent les recherches de ces sources. Par contre, les intentions ne filtreront pas des activités comme uriner ou discuter avec mon patron s'il me le demande, en dépit du fait que ces activités orientent mon attention sur autre chose que ce que j'ai planifié de faire.

Dans la réalisation d'une intention, beaucoup de stimuli et d'états somatiques « luttent » pour avoir l'attention de l'agent et le détourner de cette réalisation. Un agent peut introduire en cours de route des modifications dans l'ordre de priorité des activités qu'il s'est initialement donné sous forme de plan, tout en reste fidèle *à l'esprit du plan ou de l'intention*. On comprend que les états somatiques peuvent devenir si saillants que l'agent n'a d'autres choix que de faire quelque chose pour rétablir un état homéostatique stable s'il veut être efficace dans la réalisation de son plan ou de son intention. Pour ce qui est des stimuli externes, certains doivent être traités en priorité même si nous n'avons pas initialement pris de décision en ce sens. Lorsque nous formons une intention orientée vers le futur, nous ne spécifions pas l'ensemble des clauses d'exception qui nous autorisent à ajourner sa réalisation.

[90] Le lecteur de Bratman s'apercevra que j'utilise des termes que Bratman n'utilise pas pour exprimer ses positions. La liberté que je prends ici à cet égard reste tout de même fidèle à son propos.

Évidemment, la question des clauses incluses dans la décision de faire quelque chose dans le futur peut être une question délicate – en particulier dans un contexte juridique. Les clauses peuvent être implicites, comme dans le cas où je décide d'aller faire du ski ce week-end, *à moins que toute la neige ne soit fondue d'ici là*. Aussi, on ne me reprochera pas d'être irrésolu si j'examine d'autres alternatives de loisir voyant que la neige fond à vue d'œil, bien que personne ne m'ait entendu exprimer la clause : « à moins que toute la neige ne soit fondue d'ici là ».

Cela dit, il se peut très bien que je n'aie pas, dans certaines situations, élaboré même implicitement une clause d'exception, mais que l'on considère que mon intention inclut *virtuellement* une telle clause. Je peux décider d'aller visiter la Vallée des Rois en Égypte, mais de me raviser ensuite parce que les premiers signes d'une révolution éclosent dans ce pays. Je peux ne rien connaître de la situation politique des Égyptiens, ne pas savoir qu'un régime autoritaire y règne et que la liberté de presse est très limitée. Aussi, il ne me viendra pas à l'esprit d'inclure – ne serait-ce qu'implicitement – une clause mentionnant cette possibilité.

La plupart du temps, on identifie la teneur des clauses virtuelles *post facto*. On fait des hypothèses qui ont la forme de conditionnels contrefactuels à propos de ce que l'agent *aurait inclus* comme clause *s'il avait pu prévoir ce qui s'est effectivement passé après*. Cela ressemble beaucoup à ce que les juristes font lorsqu'ils interprètent les lois à la lumière que ce qu'ils appellent l'*esprit du législateur* (Dodd, 2007). Ils se servent de normes d'interprétation, de cas jurisprudentiels et de considérations relatives au cas particulier. Mais nous le faisons tous de manière plus ou moins intuitive à propos des autres et de nous-mêmes lorsque vient le temps de poser un jugement sur le caractère résolu d'un comportement. Imaginons, par exemple, je décide de donner le bain à mon chat au grand plaisir de la maisonnée. J'achète les produits qu'il me faut pour m'acquitter de cette besogne, mais décide de sortir avec des copains plutôt que de le faire. L'événement déclencheur est le simple fait que j'ai observé, au moment où je m'apprêtais à attraper le chat, un héron passant au-dessus de ma maison. Ma femme diagnostiquera chez moi une irrésolution et cela ne changera pas grand-chose si je lui dis : « Chérie, j'avais certes pris la décision de donner le bain au chat, mais il est dans l'esprit de cette décision que je m'abstiendrais de le faire si un héron passe au-dessus de notre maison! ».

8.1.2 Réviser une intention : une conception normative

Les caractères *stables* et *contrôlants* des intentions, plans et résolutions permettent à Holton de formuler une conception originale de la faiblesse de volonté. Pour Holton,

> « *a person exhibits weakness of will when they revise their intention, in circumstances in which they should not have revised it. This 'should' is not meant in a moral sense. Rather it is the 'should' which is generated by the norms of the skill of managing one's intentions. A person is weak willed if they revise their intention too readily.* » (1999 : 247)

La stabilité et le contrôle des intentions sont nécessaires pour atteindre des objectifs indirects ou à plus ou moins long terme. Sinon l'instabilité dynamique de nos préférences nous en détournerait trop facilement. Or, cette stabilité et ce contrôle ne doivent pas évidemment dégénérer chez les agents dans une forme d'obstination. L'obstination est une attitude irrationnelle, et les agents le savent[91]. Aussi, ils se donnent habituellement le « droit » de réviser leurs intentions (ou octroie ce « droit » aux autres) en cas d'imprévus ou de modification importante des situations. On peut réviser une intention, un plan ou une résolution pour de bonnes raisons. Si je prends la décision d'aller piqueniquer demain, mais que, la journée venue, il se mette à grêler, j'ai une bonne raison de réviser mon intention. Peut-être que mon intention comprenait une clause implicite stipulant que je ne dois pas y aller en cas de mauvais temps. Mais supposons qu'il pleuve en fait des grenouilles, j'aurais une aussi bonne raison de ne pas exécuter ma décision initiale même s'il elle ne contenait pas de clause concernant les grenouilles.

Mais on peut le faire également pour de mauvaises raisons – auquel cas le terme approprié est peut-être de « violer » une intention et non pas de la réviser. Mais qu'est-ce que cela veut dire que violer une intention ou de la réviser trop rapidement?

Selon Holton, il n'est pas rationnel de reconsidérer une décision si cette reconsidération manifeste des tendances qu'il n'est pas pour l'agent raisonnable d'avoir. Si on ajoute ceci à la proposition de Holton mentionnée plus haut, il s'ensuit que

> « […] *actors show weakness of will when they revise an intention as result of a reconsideration that they should not have performed ; that is, when their reconsideration exhibits tendencies that it is not reasonable for the agent to have.* » (1999 : 252)

[91] Nietzsche voyait plutôt dans l'obstination le signe d'un caractère fort (1886 §107).

Bien que cela semble terriblement vague, Holton soutient que cette description saisit l'essentiel de notre concept de faiblesse de volonté en mettant en relief le paramètre normatif qui le sous-tend. Rien ne nous empêche de spécifier un peu plus la nature de ces tendances. Habituellement, il est raisonnable, selon Holton et Bratman, d'avoir des tendances à reconsidérer des intentions :

(1) dans les situations qui ont changé à un point tel que les objectifs que nous poursuivions en formant ces intentions ne seront pas atteignables si nous continuons de nous y conformer;
(2) dans les situations qui présentent pour nous des aspects particulièrement déplaisants ou qui sont des sources de souffrance, mais dont nous n'avons pas tenu compte lorsque nous avons formé notre intention.

Il est aussi raisonnable d'avoir des tendances à *ne pas* reconsidérer des intentions

(3) dans les situations qui nous empêchent d'avoir l'esprit clair à propos des circonstances qui ont donné lieu à la formation de ces intentions;
(4) dans des situations où nous avons justement formé ces intentions dans le but de nous motiver à passer par-dessus une réticence à agir.

Ces tendances restent tout de même assez vagues. Mais Holton soutient que cela tient au fait que notre concept de faiblesse de volonté est vague et que la présence de vague dans l'*explanandum* correspond au vague qui se trouve dans l'*explanans*.

Cela dit, si nous avons une connaissance plus ou moins consciente de ces conditions d'arrière-plan pour établir un diagnostic de faiblesse de volonté, nous risquons de la confondre à l'occasion l'akrasie et les caprices. Il nous est tous déjà arrivé de suivre en voiture un ami qui avait choisi un restaurant, mais qui, par caprice, modifie sans cesse son choix en cours de route. Nous estimons spontanément que dans ce genre de cas, notre ami prête l'oreille à des tendances qu'il n'est pas raisonnable de considérer, même s'il croyait depuis le début que son choix initial n'était pas le meilleur choix. Les caprices peuvent rendre la coordination intersubjective exaspérante. Mais nous n'avons pas l'intuition qu'il s'agit d'un problème de faiblesse de volonté en dépit du fait que les conditions (3) et (4) ne sont pas respectées dans ce cas.

Conscient du problème, Holton distingue les intentions qui sont conçues pour être *contraires aux inclinaisons délétères* et celles qui ne sont pas conçues comme telles. Si un agent est disposé à réviser une intention qui fut – au moins en partie – conçue pour être stable face à des inclinations délétères subséquentes, alors il vivra un épisode de faiblesse de volonté. Dans le cas contraire, il se montrera seulement capricieux.

Cela dit, si l'on inclut parmi les paramètres de stabilité des intentions, plans et résolutions, des injonctions d'ignorance relative des inclinaisons délétères à venir, il est clair que la conception de Holton comporte des éléments normatifs irréductibles. Il n'y a pas pour Holton de descriptions satisfaisantes de la faiblesse de volonté qui fassent l'économie de considérations normatives. Aussi, il rejette la définition Standard comme caractérisation de la faiblesse de volonté parce qu'elle fait, entre autres, l'économie de telles considérations.

8.1.3 Des avantages à considérer la faiblesse de volonté comme la violation d'une intention

Holton expose toute une série de raisons qui militent clairement en faveur de l'idée que non seulement la définition Standard n'est pas satisfaisante pour caractériser les cas les plus communs de faiblesse de volonté, mais que sa propre conception permet d'expliquer bon nombre d'intuitions importantes que nous avons à l'égard de ce phénomène. Ces raisons sont particulièrement convaincantes et peuvent à mon avis être tout à fait complémentaires aux raisons que j'ai données dans le chapitre précédent – ou les recouper – pour rejeter la définition Standard.

8.1.3.1 Décider entre des options incommensurables ou indifférentes

Un agent peut former l'intention de s'engager dans une ligne de conduite plutôt qu'une autre sans être pour autant capable de déterminer laquelle est la meilleure parce que les options sont incommensurables. Holton mentionne l'exemple d'un jeune homme qui pendant la Seconde Guerre mondiale doit décider entre aller combattre le fascisme ou rester à la maison pour s'occuper de sa mère malade. Bien que le jeune homme ne puisse établir laquelle de ces deux options est préférable, il décide néanmoins de s'engager dans la résistance. Or, quand vient le moment d'exécuter sa décision, il se ravise par crainte d'y laisser sa vie et demeure chez lui pour s'occuper de sa mère. On peut sans problème imaginer que cet homme vit un épisode de faiblesse de volonté, même s'il n'a pas *ex*

hypothesis agit à l'encontre d'un meilleur jugement, parce que l'incommensurabilité des options ne lui a pas permis d'identifier la meilleure.

8.1.3.2 Une place pour la force de volonté

Les personnes qui emploient le terme « faiblesse » emploient aussi à l'occasion le terme « force » pour qualifier la volonté de leurs congénères ou d'eux-mêmes. Ces termes s'opposent, mais partagent le même arrière-plan conceptuel. Un agent a une volonté faible lorsqu'il révise trop rapidement une intention à la faveur de toutes sortes de considérations qu'il devrait néanmoins négliger. Mais un agent a une volonté forte lorsqu'il est capable de maintenir le cap dans la réalisation de son intention en dépit de tendances contraires plus ou moins fortes et qu'il serait irrationnel pour lui de satisfaire. Cela permet à Holton de distinguer la force de volonté de l'opiniâtreté. L'agent opiniâtre maintiendrait, comme l'agent ayant une volonté forte, le cap dans la réalisation de son intention, même s'il était approprié de réviser cette dernière. Aussi, la force de volonté peut se transformer en opiniâtreté quand l'agent devient plus ou moins insensible aux différents signaux qui militent dans le sens d'une redirection de ses efforts.

La distinction entre la force de volonté et l'opiniâtreté est quelque peu vague. Il existe beaucoup de cas limites pour lesquels les observateurs ne s'entendront pas sur le diagnostic. Cela n'implique évidemment pas que la distinction soit erronée, mais seulement que, étant donné le caractère vague des deux notions, l'explication de Holton comportera des aspects vagues.

Avec la conception classique de la faiblesse de volonté, il devient difficile d'offrir une explication de la faiblesse et de la force de volonté qui partage un même schème. On contraste, dans la conception classique, la faiblesse de volonté avec le contrôle de soi. Le contrôle de soi est aussi présent dans la force de volonté que dans l'opiniâtreté et ne permet donc pas de distinguer les deux. D'ailleurs, le problème que peut représenter pour un agent le fait d'avoir un contrôle de soi trop marqué réside dans le manque de spontanéité qu'il implique. Or, comme le soutient Holton, le manque de spontanéité est très différent de l'opiniâtreté.

8.1.3.3 L'absence de conflits internes

Pour la conception classique, un agent qui vit un épisode de faiblesse de volonté vit en son for intérieur nécessairement un conflit. Si un agent akratique vit un épisode de faiblesse de volonté, alors il doit s'engager consciemment dans une ligne de conduite sous-optimale en dépit du fait qu'il pense connaître une ligne de conduite optimale. Lorsque nous faisons un diagnostic à propos de nous-mêmes, nous nous servons bien souvent de la tension interne que nous éprouvons comme indice que nous vivons un épisode de faiblesse de volonté.

Or, nous ne faisons pas toujours la même chose lorsque vient le temps de faire un diagnostic en 3^{e} personne. Nous nous contentons de connaître la décision initiale que l'agent a prise à l'égard de ses activités futures, s'il l'a exécuté ou non, et dans le l'éventualité où ce n'est pas le cas, nous essayons d'identifier de bonnes raisons dont il aurait pu apprendre l'existence et sur la base desquelles il aurait décidé de réviser son intention initiale. Nous n'avons pas besoin de faire des hypothèses sur la teneur d'un supposé conflit interne que vit l'agent parce qu'il se peut très bien qu'il n'en vive tout simplement pas. L'agent peut avoir plus ou moins subi une modification de son état d'esprit de manière à ce qu'il considère maintenant moins importantes les raisons qu'il avait au moment où il a formé son intention. La consommation d'alcool diminue typiquement l'aversion au risque. Un agent qui a pris la décision de revenir chez lui en taxi s'il consomme plus d'un verre de bière dans la soirée sera disposé à réviser son intention après intoxication parce qu'il aurait – entre autres – une aversion au risque beaucoup moins élevé, ou alors inexistant. Mais en dépit du fait que son esprit a changé et qu'il ne semble pas éprouver le moindre conflit interne, on sera justifié de diagnostiquer dans son cas un épisode de faiblesse de volonté.

Les diagnostics en 3^{e} personne sont plus fiables parce qu'ils font fi des mécanismes de réduction de la dissonance cognitive à l'œuvre chez les agents akratiques. Ces mécanismes sont souvent responsables des sous-diagnostics que ces agents font régulièrement à propos d'eux-mêmes.

8.1.3.4 Akrasie tout court et stricte akrasie: des sous-catégories en partie exclusives

L'analyse de Holton lui permet de distinguer les cas de stricte akrasie et les cas d'akrasie – auxquels cas il réserve plutôt le terme « faiblesse de volonté ». Cela lui permet de dire que

sa conception est complémentaire de la conception classique et ne vise pas vraiment à la remplacer.

Conformément à la définition Standard nous vivons un épisode de stricte akrasie lorsque nous agissons à l'encontre de notre meilleur jugement. Aussi, je peux, toutes choses considérées, juger préférable de ne pas manger de viande – entre autres pour des raisons morales – tout en continuant d'en manger. Pour Holton, il s'agit d'un cas de stricte akrasie, mais pas d'akrasie (ou faiblesse de volonté). Aussi, ce cas est différent de celui où je déciderais *en plus* d'arrêter de manger de la viande, mais échouerait à exécuter ma décision. Dans ce cas uniquement on pourra faire un diagnostic d'akrasie (ou de faiblesse de volonté) à propos de mon comportement.

Également, je peux me conformer à mon meilleur jugement en révisant une décision irrationnelle que j'ai prise précédemment. Ce faisant, je peux vivre un épisode d'akrasie sans pour autant que mon comportement relève de la stricte akrasie. Holton mentionne le cas hypothétique d'une femme non mariée à l'époque victorienne qui décide de s'embarquer dans une affaire amoureuse qu'elle sait pourtant potentiellement désastreuse. De plus, cette femme considère que cette relation est moralement condamnable et va sûrement la laisser enceinte. Néanmoins, elle décide d'aller de l'avant. Mais au dernier moment, tout juste avant de communiquer un geste d'engagement avec l'élu de son cœur, elle perd son courage et se désiste.

8.1.3.5 Compatibilité avec des procédures décisionnelles variées

Bien que Holton ne mentionne pas ce fait, sa conception comporte un avantage que j'estime particulièrement important. Elle n'est pas tributaire d'une conception unique et étriquée des procédures décisionnelles rationnelles comme la conception classique. Les agents peuvent utiliser toutes sortes d'heuristiques dans la formation de leur jugement ou la prise de leur décision. Ils peuvent faire des choix par défaut, satisfier, éliminer les extrêmes, choisir aléatoirement, faire un choix en fonction d'une seule raison discriminante, se servir d'un proxy affectif, etc. Du moment qu'ils ont formé l'intention de faire quelque chose dans le futur en dépit des inclinations délétères qui sont susceptibles de les faire dévier de leur objectif initial, ils peuvent vivre des épisodes de faiblesse de volonté.

Ceci rend la conception de Holton particulièrement attrayante. Elle laisse ouverte la question de savoir comment les agents décident. C'est une question empirique et les philosophes ne sont certainement pas les chercheurs les mieux positionnés pour y répondre. Par contre, ils peuvent spécifier assez bien les conditions normatives de révision rationnelle des décisions antérieures.

8.1.3.6 La stigmatisation des épisodes de faiblesse de volonté

Holton soutient que sa conception nous permet de comprendre pourquoi on stigmatise spontanément les agents akratiques. Un ami ou un parent peut soutenir devant nous qu'il serait préférable pour lui de s'abstenir de fumer, de manger des friandises ou de sortir du lit en retard le matin. On sera d'accord avec lui et l'on ajoutera que, s'il était continent, il s'abstiendrait d'agir de la sorte. On le stigmatisera cependant s'il décide, par exemple, d'arrêter d'agir de la sorte à partir du 1er janvier et qu'on l'observe dans ses anciennes habitudes le 2 janvier, et non à partir du moment où il nous fait part de son jugement selon lequel il serait préférable de ne pas agir de la sorte.

8.2 Quelques difficultés pour la théorie de Holton

Comme je viens de le mentionner, la conception de Holton comporte beaucoup d'avantages : elle est simple, claire, élégante et, en plus de s'accorder assez bien avec nombre de nos intuitions, elle constitue certainement un complément utile à la conception Standard. C'est ce qui explique pourquoi, je pense, elle est aujourd'hui très influente. Pourtant, elle a son lot de difficultés.

Premièrement, il est souvent difficile de distinguer les cas où un agent révise ses intentions à la faveur d'une modification des circonstances des cas où il les adapte à ces nouvelles circonstances. De même, il semble que Holton néglige la distinction pourtant intuitive entre les cas où les agents révisent effectivement une intention et les cas où ils ne font qu'y désobéir. Prendre en compte ces éléments a pour résultat d'évacuer les considérations normatives si essentielles aux yeux de Holton.

Deuxièmement, la notion d'intention, telle que décrite par Bratman et reprise par Holton, prête le flanc à plusieurs critiques. D'aucuns estiment que les intentions sont d'étranges

entités psychologiques, et seuls certains philosophes croient en leur existence. Il est par conséquent hasardeux d'en faire le concept central d'une théorie quelle qu'elle soit.

Troisièmement, il y a beaucoup de cas putatifs de faiblesse de volonté qui ne correspondent pas au portrait que fait Holton, pas plus qu'ils ne correspondent au portrait de la conception Standard, notamment les cas de négligence ou d'apathie.

8.2.1 Réviser et interpréter une intention, et lui désobéir

Le philosophe Dylan Dodd a formulé une critique intéressante de la conception de Holton. Il remet en question l'hypothèse que la faiblesse de volonté advient dans des situations où les agents procèdent à la révision de leurs intentions. Pour Dodd, les révisions ou les annulations d'intention, dont parlent Bratman et Holton, ne correspondent pas à la plupart des cas réels. Si je forme l'intention d'aller piqueniquer demain, et que le lendemain il pleut, je ne réviserai pas mon intention si je m'abstiens d'y aller. Je n'agirai pas à l'encontre de mon intention parce qu'elle comporte notamment des clauses implicites concernant la météo. Mais, comme je l'ai mentionné plus haut, nos intentions ne comportent pas des clauses pour toutes les situations possibles. C'est ici que l'interprétation entre en jeu. Je peux interpréter une décision antérieure et juger que mon action actuelle se conforme à son esprit. Si j'ai raison dans ces conditions, il se peut que mon diagnostic initial de faiblesse de volonté soit erroné. Par exemple, je peux décider que je dois travailler au bureau toute la soirée, mais être témoin dans le courant de cette soirée de l'agression très violente d'un étudiant par des individus déguisés en clown. Aussi, je décide de lui venir en aide et passe, par conséquent, le reste de la soirée au poste de police. Au moment où j'ai formé mon intention, je n'avais évidemment pas à l'esprit ce genre de condition d'exclusion. Mais je peux raisonnablement considérer que mon action reste conforme à l'esprit de mon intention initiale.

Si cela est correct, on peut alors estimer avec Dodd que la principale « charge » normative de l'agent dans une situation de la sorte n'est pas de déterminer s'il est rationnellement justifié de réviser son intention, mais de déterminer s'il est rationnellement justifié d'avoir telle ou telle interprétation de son intention initiale.

Dodd ne nie pas qu'on puisse réviser une intention, mais soutient que Holton confond souvent les deux et que cela l'amène à élaborer une conception inexacte. Je peux, par

exemple, former l'intention de faire du jogging pendant 15 minutes à 6 heures du matin toute la semaine. Mais, après quelques semaines, je peux reconsidérer ma décision et en arriver à la conclusion que cela ne vaut pas vraiment la peine pour le peu de bénéfices que cela représente pour ma santé. Il est clair ici que je prends la décision de réviser mon intention. Par contre, on peut tout à fait agir à l'encontre d'une intention comme une politique individuelle *sans l'annuler pour autant*. La différence entre réviser ou annuler une intention et la faiblesse de volonté est analogue à la différence qu'il y a entre l'abrogation d'une loi gouvernementale et la désobéissance des citoyens à l'égard d'une loi encore en vigueur. Aussi, pour Dodd, un agent vit un épisode de faiblesse de la volonté si et seulement s'il *viole* une intention ou *désobéit* à un plan ou une résolution et non s'il l'a révise.

Il n'est toutefois pas toujours aisé de distinguer ces deux types de cas. Les intentions ne sont pas toujours clairement explicites, et l'agent n'en a pas toujours conscience. La présence de regrets subséquents peut être un indice utile pour faire un diagnostic de désobéissance. Si je prends la résolution d'arrêter de fumer à partir de demain, mais que vous m'observiez succomber à la tentation une semaine plus tard, vous allez suspecter un épisode de faiblesse de volonté. Si je regrette amèrement mon écart de conduite ensuite, vous aurez alors une bonne raison de croire que je n'ai pas révisé ma résolution avant de fumer, que celle-ci est toujours en vigueur, et non que j'ai formé une nouvelle résolution. Cela n'empêche pas cependant qu'au moment de ma rechute j'ai pu tenter, pour moi-même et pour les autres, de justifier rationnellement le fait de m'allumer une cigarette, ce qui laisse penser que j'ai révisé ma résolution. Mais comme la duperie de soi est un phénomène souvent corrélatif des épisodes de faiblesse de volonté, on doit rechercher d'autres indices pour établir un diagnostic plus robuste. Aussi, comme le soutient Dodd,

> « […] *whether an agent A has a certain policy at t cannot always be determined by consulting a synchronic description of A at t, but can often only be determined by a diachronic description of A. Whether he really did cancel the old policy will be shown by how he acts and feels later. Subsequent guilt feeling, promises never to do that again, and the like are all again that the policy was still in effect and broken. A lack of such things combined with a stable disposition to affirm the one has decided not to quit smoking after all would show that one really did decide to cancel the old «* Stop smoking! *» policy. Regardless, if an agent really did cancel the old policy of refraining from smoking and only then smoked, his smoking wouldn't be a case of weakness of will. Because the policy was canceled, there would be no policy the agent had which he would be breaking. However, if the subsequent behavior and affirmations of the agent reveal he never did cancel that old policy, in spite of what he might have said to the contrary as was giving into temptation, then he would have*

> *been acting against a policy he still had by smoking. Then he would have been weak-willed.* » (2007 : 54)

Il n'est évidemment pas toujours aisé de distinguer les cas de révision des cas de violation d'intention. Mais dans la mesure où l'on peut le faire, il semble qu'on n'ait plus besoin de faire référence à des considérations normatives constitutives, comme celles que Holton estime être nécessaires pour une caractérisation adéquate de la faiblesse de volonté et son diagnostic. Pour Dodd, les considérations normatives *peuvent* intervenir, mais seulement pour diriger les interprétations des intentions initiales. J'ajouterais que comme l'interprétation d'une décision n'est nécessaire que dans les cas ambigus ou peu clairs, cela nous amène à affaiblir grandement la dimension normative que Holton considère pourtant centrale pour toute bonne conception de la faiblesse de volonté.

8.2.2 Des problèmes avec la notion d'intention?

La nécessité de poser des intentions comme entité mentale pour expliquer le comportement des agents a suscité beaucoup de commentaires et de critiques dans le milieu philosophique[92]. Il n'est pas dans mon intention ici de relater l'ensemble de ces critiques. La question des intentions est complexe et recouvre des phénomènes hétérogènes.

On peut faire appel aux intentions et mettre en relief le rôle qu'elles jouent dans la rationalisation que les agents font de leurs propres comportements dans une situation où ils doivent rendre des comptes à autrui (Thomson, 2008). Ce genre de rationalisation permet entre autres d'identifier pour des observateurs les objectifs principaux, secondaires ainsi que les effets collatéraux voulus ou non engendrés par la poursuite de ces objectifs. C'est ce qui arrive lorsqu'on demande à quelqu'un de nous expliquer pourquoi il fait *A*. Il peut répondre qu'il fait *A* parce qu'il a l'intention de faire *B*, mais n'a pas l'intention de faire *C* même si *C* découlerait probablement de la réalisation de *B*, etc. On parlera d'*avoir l'intention de* X dans ce genre de situation. L'expression serait utilisée en conjonction avec d'autres (vouloir que, savoir que, etc.) pour établir une sorte de carte de la structure motivationnelle que représente pour l'agent son environnement.

[92] Pour un aperçu, voir l'entrée « Intention » de Kieran Setiya dans *The Stanford Encyclopedia of Philosophy*.

On peut également mettre en relief le rôle qu'elles peuvent jouer pour établir des cas de négligence criminelle. Dans un cas de négligence criminelle, l'agent n'a pas l'intention de faire quelque chose qui cause du tort à autrui. Il ne fait qu'ignorer certaines conséquences néfastes que son action produit sur autrui (Davis : 1979). Il ignore ces conséquences alors qu'il aurait dû prendre des dispositions minimales pour les connaître avant d'accomplir son action. On essaiera de savoir si le prévenu a fait *X* intentionnellement dans ce genre de situation.

Ces usages de la notion d'intention sont relativement non problématiques. Par contre, faire des intentions des états mentaux réalisés dans le cerveau des agents et distincts de leurs préférences apparaît à certains comme étant un pas de trop. Pour Heath, par exemple, les intentions jouent le même rôle que les apparences, les données sensorielles ou les qualia que les philosophes ont de manière erronée posés entre les stimuli qui affectent les sujets connaissants et les rapports qu'ils font à leur propos. Pour expliquer le fait que les sujets connaissants peuvent fournir des rapports erronés sur les objets qu'ils observent pourtant directement, les philosophes ont cru bon de postuler des entités mentales situées entre ces objets et les croyances qu'ils entretiennent à leur égard. Les croyances ne portent pas directement sur le monde extérieur, mais sur des apparences que le monde cause en nous. Cette astuce théorique permet, soutient Heath, d'établir, d'une part, un lien incorrigible entre les apparences et les croyances, et, d'autre part, la possibilité que nos croyances soient systématiquement fausses (2008 : 160-165). Le premier point semble nécessaire parce que, de toutes les affirmations que nous faisons à propos des objets dont nous faisons l'expérience et qui ont la forme « *x* est *P* », où *P* représente une qualité observable, on peut inférer que « *x* me semble être *P* ». Or, *x* peut ne peut pas être *P,* mais me sembler tout de même être *P*. Le second point n'est pas nécessaire, mais s'inscrit bien dans une tradition épistémologique qui aime les débats sur le solipsisme et le fondement de la connaissance.

Selon Heath, postuler des apparences dans l'esprit des sujets relève d'une confusion sur le sens des termes. Lorsqu'on utilise des expressions comme « cela me semble bleu », « cela m'apparaît être une gamme ascendante » ou encore « je sens que c'est pointu », c'est qu'on

veut attribuer à nos rapports un statut épistémique[93] plus faible que quand on utilise des expressions comme « c'est bleu », « c'est une gamme ascendante » ou « c'est pointu ». On suggère à notre interlocuteur que notre rapport est plus ou moins fiable et qu'il devrait le prendre ou non avec un grain de sel. Pour ce qui est du risque d'erreur systématique, Heath se sert de l'argument de Sellars pour rejeter cette possibilité. L'argument est simple, si nous croyons systématiquement que ce qui, par exemple, est dans la réalité vert, mais est en fait transparent, c'est que le concept que nous utilisons n'est pas celui de transparence, mais celui de vert.

À propos des intentions, Heath avance que les philosophes qui les posent comme des entités mentales sont coupables de la même confusion que ceux qui ont postulé l'existence des apparences. Les intentions semblent utiles pour rendre compte de la possibilité de l'échec. Aussi, les agents ne décideraient pas de réaliser une action, mais décideraient plutôt de former une intention de réaliser une action. L'échec d'un agent proviendrait du fait que son intention n'a pas réussi à causer l'action, et c'est ce qui expliquerait entre autres les épisodes de faiblesse de volonté.

Pour Heath, ce genre d'hypothèse relève également d'une confusion produite par une mauvaise compréhension du statut épistémique des énoncés d'intention. Lorsque nous disons à quelqu'un que nous avons l'intention de nous rendre chez un ami, nous utilisons une expression qui marque le caractère incertain du dénouement futur de notre action et que notre interlocuteur ne doit pas trop compter dessus. De même, si nous échouons systématiquement à réaliser nos intentions, c'est probablement parce que nous avons en fait décidé de faire autre chose.

Cette confusion serait responsable, selon Heath, du caractère étrange de la fameuse énigme de la toxine de Kavka (1983). Gregory Kavka a élaboré le scénario hypothétique suivant : imaginons qu'un excentrique millionnaire vous offre l'opportunité de gagner une très grosse somme d'argent si vous formez l'intention de boire un liquide toxique qui vous rendra affreusement malade pendant une journée, mais n'entraînera aucune séquelle. Par

[93] Heath parle plutôt ici de statut déontique. Il s'inspire beaucoup de la sémantique de Brandom qui soutient entre autres que les contenus des désirs et des croyances sont en partie déterminés par des règles d'autorisation et d'engagement qui cautionnent les « entrées » et « sorties » du langage. Mais, il n'est pas nécessaire d'être au fait de la théorie de Brandom pour comprendre ici la position de Heath. C'est pourquoi, je pense, on peut utiliser la notion de statut épistémique tout en restant conforme au propos de Heath.

contre, vous n'avez pas à boire le liquide pour être effectivement récompensé, seulement former l'intention de le faire. Pour s'en assurer, on vous mettra dans un scanneur, conçu par le Docteur X, qui est capable de détecter des contenus de pensée. Pouvez-vous former l'intention de boire la toxine?

Si l'on adhère à une théorie réaliste des intentions, le dilemme semble insoluble. D'une part, nous formerions des intentions parce que c'est dans notre intérêt de le faire, et, d'autre part, former ici l'intention de boire le liquide toxique va clairement à l'encontre de notre intérêt. Ce genre de cas met en relief le fait qu'il semble rationnel de former une intention de faire quelque chose, mais irrationnel de faire cette chose. Heath soutient que dans ce genre de cas il n'est pas psychologiquement impossible de former une intention de la sorte, mais logiquement impossible en vertu des règles qui encadrent l'usage des expressions « intention » et « décision »[94]. Pour Heath, la question de savoir si un agent a l'intention de faire quelque chose ne relève pas d'une question de fait – du moins pas d'une question qu'on peut résoudre avec un scan du cerveau. Attribuer une intention particulière consisterait seulement à assigner un statut doxastiquement plus faible à l'agent que de lui attribuer une décision.

On peut être en accord avec certains des arguments de Heath sans pour autant souscrire à son arrière-plan théorique et conceptuel[95]. L'idée que nous devions former des intentions pour agir et non seulement prendre des décisions est une hypothèse contestable. Cela devient particulièrement évident quand les décisions visent l'obtention de quelque chose tandis que leurs intentions corrélatives visent autre chose. Mais je pense que cela ne rend pas caduque la conception de Holton sur la nature de la faiblesse de volonté. En fait, je ne vois pas pourquoi on pourrait s'empêcher d'identifier les intentions, plans et résolutions, avec des décisions orientées vers le futur, *mais qui comportent une note sur le caractère incertain de leur réalisation* ou *même sur le caractère incertain de l'agent lui-même à l'égard de leur réalisation*. Cela nous permettrait de préserver l'essentiel de la conception de Holton et de son arrière-plan théorique. Par exemple, Holton emprunte l'idée de Bratman de distinguer les intentions des préférences. Or, si les intentions sont des

[94] Le problème n'est pas qu'un agent ne puisse essayer suffisamment fort de former ce genre d'intention, mais il est analogue à celui qui consisterait pour un agent à essayer de croire qu'une prémisse est fausse et que la conclusion qu'elle entraîne est vraie, alors qu'il accepte la validité de l'argument (*Ibid.* : 165).
[95] Ce qui est mon cas.

décisions, l'idée de les distinguer ne pose pas de problèmes, même si les deux sont conçues comme des états ou événements mentaux[96].

On pourrait objecter cependant que les intentions ainsi conçues ne peuvent pas jouer le rôle contrôlant que Holton et Bratman veulent pourtant leur faire jouer. Je peux prendre une décision pour l'avenir, mais rien ne m'oblige à respecter cette décision si cette décision ne concerne évidemment que moi[97]. Il n'y a pas d'obligations, qu'elles soient de nature morale ou non, envers soi-même[98]. Alors comment peut-on considérer que les intentions, plans et résolutions ont une dimension contraignante pour l'agent qui les forme? Une réponse consisterait à dire que les agents peuvent se contraindre eux-mêmes au moyen d'un engagement préalable, comme l'a décrit Elster (1979 : 101-114). Je peux contraindre mon moi futur en modifiant présentement l'étendue des options disponibles pour lui. Par exemple, je peux décider de ne sortir qu'avec la quantité d'argent liquide nécessaire pour faire mes achats et laisser chez moi mes cartes de crédit et de débit. Aussi, s'il me vient l'envie de faire des dépenses inconsidérées, je serai limité par les moyens financiers dont je dispose à ce moment. Or, le problème avec cette réponse est que seulement certaines intentions prennent la forme d'engagements préalables. La plupart du temps, les agents

[96] Pour Holton, former une intention, un plan ou résolution est une sorte d'acte mental dont les agents font l'expérience et dont ils ont l'impression qu'elle relève de leur volonté libre. Mais cette expérience phénoménale n'est pas seulement présente pour les intentions orientées vers le futur. Nous avons tous une expérience de la formation d'intention dont nous souhaitons la réalisation *hic et nunc*. Nous ne décidons pas nécessairement pour l'avenir, mais également pour le présent. Aussi, on peut être tenté d'identifier les intentions orientées vers le présent avec les actions qu'entreprend présentement un agent (Pettit, 2010). Cela reste d'ailleurs dans l'esprit de la conception aristotélicienne de la rationalité pratique. Pour Aristote, les actions d'un agent suivent immédiatement la délibération et elles représentent en ce sens sa conclusion. Mais identifier les intentions avec des actions revient à identifier les décisions exécutables *hic et nunc* avec les actions, ce qui est une erreur.

[97] Le problème du caractère contraignant qu'une intention représente pour soi-même est au cœur de débat actuel sur la nature des normes sociales et la possibilité de la coopération (Gautier, 1997). Les chercheurs s'entendent pour dire que la coopération requiert dans beaucoup de situations que les agents soient en mesure de punir les autres en cas de choix opportuniste de leur part. Or, la punition est coûteuse et les agents ont besoin de faire des menaces – explicites ou implicites – pour motiver les autres à s'abstenir d'adopter des comportements opportunistes. Pour que les menaces soient efficaces, un agent doit donc être disposé à exécuter une punition même s'il est rationnel pour lui de ne pas l'exécuter au moment venu parce que cela consisterait à ajouter des coûts supplémentaires (associés à punir et à prendre le risque d'une escalade de violence) à des coûts qu'il a déjà payés (associés à la défection). Aussi, faire des menaces crédibles ressemble en quelque sorte à l'énigme de la toxine. Il est rationnel de former l'intention de punir autrui en cas de non-coopération de sa part, mais pas d'exécuter cette intention advenant un comportement non coopératif de sa part.

[98] Ruwen Ogien a formulé une série d'objections à l'opinion qui veut qu'on puisse avoir des obligations envers soi-même, du moins des devoirs stricts envers soi-même (2007 : 33-57). L'objection la plus forte s'appuie sur une analogie avec les devoirs envers les autres. Pour faire court, une obligation que j'ai envers autrui a une force contraignante que si je ne m'en acquitte pas *et* qu'autrui ne la résilie pas de son plein gré. La décision de résilier une obligation ne peut pas venir de moi, mais de celui qui est le bénéficiaire de cette obligation. Mais comme je peux toujours résilier un contrat que j'établis avec moi-même, il s'ensuit qu'il ne peut être contraignant.

La portée de cette objection est cependant limitée. Holton soutient que les engagements pratiques engendrés par les intentions ne sont pas de nature morale. Mais il est difficile d'en saisir le caractère engageant dans un contexte intrasubjectif.

prennent des décisions pour l'avenir sans modifier la structure des opportunités disponibles pour eux. Comment peut-on alors parler de contrôle quand les agents restent libres de faire ce qu'ils préfèrent en dépit du fait qu'ils ont pris une décision pour l'avenir?

Je pense qu'on peut accepter l'idée que les décisions pour l'avenir comportent des aspects « contrôlants » – contrairement aux jugements préférentiels ou aux jugements inconditionnels – sans pour autant que ces décisions créent des contraintes extérieures. Les décisions peuvent enclencher des séries d'inhibitions virtuelles ou réelles des reconsidérations futures ou simplement orienter l'attention. Les personnes qui éprouvent des difficultés à enclencher de telles inhibitions sont typiquement capricieuses ou ont du mal à garder leur attention sur leurs objectifs principaux.

D'ailleurs, le contrôle qu'un agent peut exercer sur lui-même – ou sur un autre – ne résulte pas nécessairement en une restriction des options disponibles, mais peut résulter dans la modification du caractère désirable de ces options. Les lois d'un pays ne rendent pas nécessairement physiquement non réalisables des options qu'ils veulent interdire, mais ils rendent coûteuse leur réalisation au moyen de dispositifs légaux comme les amendes. De la même manière, le fait que j'ai pris une décision antérieure à l'égard d'une ligne de conduite particulière, comme une politique personnelle, peut rendre les autres lignes de conduite moins attrayantes. Ceci peut s'expliquer par le fait que dévier de ma ligne de conduite fera de moi une personne irrésolue, ce qui est incompatible avec l'image que je me fais de moi-même. Mais cela peut également s'expliquer par le fait que l'exécution de ma décision initiale a nécessité un « investissement » en temps et en énergie dont je devrai assumer la perte si je m'engage dans une autre ligne de conduite. Il est vrai que, dans ce cas, c'est l'exécution d'une décision qui rend son abrogation plus coûteuse et non la décision elle-même. Mais comme toutes les décisions appellent une exécution, elles produisent du moins potentiellement ce genre de facteur de contrôle.

Je suis conscient que je n'offre ici qu'une réponse sommaire à l'objection qui veut que les décisions que l'on prend pour l'avenir ne puissent avoir la dimension contrôlante postulée par Holton et Bratman. Bien que chacun accepte l'idée que les plans et résolutions servent de guide pour l'action, il est plus difficile d'accepter l'idée que ces guides créent des contraintes internes dans la poursuite d'objectifs divergents.

8.2.3 Des cas qui restent inexpliqués

Les analyses conceptuelles comme celle de Holton souffrent toutes du même défaut. En essayant de formuler un critère identifiant les conditions nécessaires et suffisantes d'un phénomène aussi complexe et varié dans ses manifestations, Holton en restreint trop la classe des cas putatifs. Je ne critique pas ici sa décision de distinguer les cas de faiblesse de volonté (ou akrasie) des cas de stricte akrasie. Il peut être judicieux de scinder l'analyse d'un phénomène qui comporte des aspects hétérogènes en des chapitres différents. Mais, à mon avis, l'analyse de Holton, bien qu'ingénieuse, est quelque peu insatisfaisante, et cela, pour deux raisons.

La première est que Holton ne nous montre pas ce que la faiblesse de volonté a de commun avec la stricte akrasie. Bien que regroupant des éléments hétérogènes, nous avons tout de même l'intuition forte que l'irrésolution, certaines inconsistances dynamiques de nos préférences, les cas de désalignement des mécanismes de motivation et d'évaluation, et même les cas de stricte akrasie, comportent des traits communs fondamentaux qui nous permettent de les classer dans la catégorie des choix irrationnels. Holton ne fait que nous donner des indications, que d'aucuns considéreraient trop vagues, sur certaines normes de révision que les agents doivent respecter sans quoi ils s'exposent à des épisodes de faiblesse de volonté. Par contre, je pense que Holton a raison de soutenir qu'une conception purement descriptive est inadéquate. Si l'on accepte l'idée que la faiblesse de volonté est le paradigme de l'irrationalité pratique, alors on doit d'une manière ou d'une autre inclure des considérations normatives dans sa caractérisation. En s'abstenant de le faire – comme c'est le cas pour la définition Standard – on « découvre » alors une panoplie de cas de faiblesse de volonté qu'il est rationnel d'avoir[99].

La seconde raison pour laquelle la conception de Holton est insatisfaisante est qu'elle ne porte que sur la *non-observance* d'une autoprescription antérieure. Or, nous avons tous l'intuition que même les personnes qui ne prennent pas de résolutions ou ne s'engagent pas à faire des activités dans l'avenir peuvent vivre des épisodes de faiblesse de volonté ou présenter des traits de caractère typiques d'agents dont la volonté est faible. Donc, si la conception de Holton est correcte, un agent ne s'exposerait à vivre un épisode de faiblesse

[99] Ces cas furent, entre autres, identifiés par Ogien (2003), Audi (1990) et McIntyre (1990).

de volonté que dans la mesure où il fait des choix pour l'avenir. Cela exclut donc l'hypothèse qu'une personne vivant dans la complète indolence ou oisiveté éprouve des problèmes de faiblesse de volonté. Mais cela exclut également le fait que des animaux non humains vivent de tels épisodes[100]. D'ailleurs, si l'on accepte l'idée que l'impatience et la procrastination relèvent de la faiblesse de volonté, alors on voit que l'analyse de Holton est trop restrictive. Il y a beaucoup de cas d'impatience et de procrastination qu'on ne peut pas traiter comme des cas d'irrésolution. Reste l'option de les traiter comme des cas de stricte akrasie. Mais comme je l'ai montré dans le chapitre précédent, très peu de jugements sont issus d'une délibération. Les agents utilisent dans la plupart des situations des règles décisionnelles plus simples et adaptées, même quand ils se montrent impatients ou procrastinateurs.

Du reste, la conception de Holton ne nous dit rien du caractère akratique mais seulement des épisodes comme tels de faiblesse de volonté. L'indolence, la paresse et l'aboulie comme trait de caractère ne sont pas explicables dans la conception de Holton parce que les agents qui ont ces traits ne prennent pas nécessairement des résolutions pour l'avenir – certains vivent au jour le jour et il est difficile de justifier l'idée que leur volonté est faible si on adhère à la conception de Holton.

8.3 Conclusion

Ce qui est intéressant avec la théorie de Holton est qu'elle fait partie des trop rares théories qui portent sur la forme diachronique de la faiblesse de volonté et non sur la stricte akrasie. Pour avoir une théorie aussi élaborée, on doit remonter à Platon et à Aristote. La théorie présente cependant certaines difficultés. Mais ces difficultés ne sont pas insurmontables. Elles ne nécessitent que quelques précisions et aménagements sommaires.

Par exemple, il peut être conceptuellement difficile de distinguer les violations d'intention des simples révisions, et l'on peut critiquer l'idée qu'introduire dans la théorie un élément normatif prétendument constitutif du phénomène. Or, on faire reposer la distinction sur nos intuitions communes pour les cas clairs et sur l'interprétation de l'esprit de l'intention pour

[100] Si, bien sûr, on accepte l'idée que les animaux – en raison peut-être de l'absence de compétences langagières référentielles – ne peuvent avoir de désirs indexés au temps comme le suggère De Sousa (1987 : 205-233).

les cas ambiguës. À ce moment, l'élément normatif que Holton croit détecter dans le phénomène est en fait une norme d'interprétation. Aussi, la notion d'intention comme événement ou état mental distincte des croyances et désirs devient conceptuellement irréprochable si on l'assimile à celle de décision orientée vers l'avenir.

Je ne sais pas s'il convient ou non de parler ici de difficulté, mais la théorie de Holton réduit les cas de faiblesse de volonté non strictement akratique à des cas d'irrésolution. Pourtant, certaines formes d'apathie, de négligence, d'imprudence ou d'insouciance relèvent intuitivement de cas de faiblesse de volonté. On ne peut pas décrire ces cas comme des violations d'intention ou de résolution. De même, certains comportements impatients et impulsifs ne se laissent pas aisément concevoir comme des cas d'irrésolution. Une personne qui dépense trop et de manière systématique n'agit pas nécessairement à l'encontre d'une décision budgétaire qu'elle se serait donnée, même dans les cas où elle regrette amèrement ces actes de consommation.

Sans que cela soit une critique, nous devons constater que Holton ne fournit qu'un cadre conceptuel. Il n'explique pas vraiment pourquoi les agents sont tantôt résolus, tantôt irrésolus. Sa conception ne s'appuie pas sur une théorie générale de la motivation qui lui donnerait plus de profondeur.

J'exposerai dans le prochain chapitre les éléments d'une théorie plus générale de la motivation qui se veut assez puissante et empiriquement robuste pour expliquer autant les cas de stricte akrasie que les cas d'akrasie tout court. Cette théorie n'est cependant pas une alternative aux conceptions discutées jusqu'ici, mais une sorte de complément conceptuel robuste sur la base duquel on peut orienter la discussion du problème de la faiblesse de volonté.

9

La théorie du choix intertemporelle (1re partie) : distribution des conséquences des décisions akratiques

Some Choices we live not only once but a thousand time over, remembering them for the rest of our life

–Richard Bach

9.1 Vers une théorie plus générale

La conception Standard et la théorie de Holton sont insatisfaisantes pour les raisons que j'ai mentionnées. Parce qu'elle repose sur une conception erronée et idéalisée des processus décisionnels et de formation des jugements, la conception classique aboutit à un surdiagnostic des épisodes de faiblesse de volonté. Même si la conception de Holton est compatible avec une variété de processus décisionnels, elle aboutit quant à elle à un sous-diagnostic de ces épisodes parce qu'elle traite la faiblesse de volonté exclusivement comme un problème de non-observance de résolution. L'analyse conceptuelle de Holton reste néanmoins ingénieuse et peut certainement constituer un chapitre important d'une théorie plus générale du phénomène.

Aussi, je propose maintenant d'aborder les éléments conceptuels fondamentaux d'une théorie plus générale qui permet de rendre compte autant des intuitions que nous avons sur le phénomène que des données expérimentales mises en relief ces dernières années par les psychologues et les économistes. Ces éléments devraient nous permettre de fournir des éléments de réponse satisfaisants à toute une panoplie de questions relatives au phénomène, ainsi que d'en soulever de nouvelles qui sont susceptibles de recevoir un traitement empirique. Une des questions les plus importantes, mais étrangement rarement abordée dans la littérature philosophique, est celle de savoir pourquoi les décisions strictement akratiques (faiblesse de volonté synchronique) *et* akratiques tout court (faiblesse de volonté diachronique) concernent *les mêmes activités* ou portent sur *les mêmes événements*. *Grosso modo*, notre volonté s'affaiblit lorsque vient le temps d'accomplir certaines activités ou de

subir certains traitements, malgré le fait que nous ayons ou non préalablement formé un jugement selon lequel qu'il serait approprié de les faire ou de les subir.

D'ailleurs, en utilisant ces éléments, nous devrions être plus facilement en mesure de savoir pourquoi nous vivons des épisodes de faiblesse de volonté dans certaines situations, mais pas dans d'autres, et pourquoi nous sommes plus susceptibles de procrastiner, lorsque vient le temps d'accomplir des besognes, et de s'adonner à des « vices ». Ces éléments devraient pouvoir nous orienter dans l'examen des facteurs endogènes et exogènes de changement des préférences et désirs qui produisent des épisodes de faiblesse de volonté, mais également des facteurs de contrôle de soi et diverses méthodes pour accentuer ce contrôle.

Ils nous permettent également de voir pourquoi la faiblesse de volonté est un paradigme de l'irrationalité, ce qui est beaucoup plus difficile à faire dans le cadre de la conception Standard et même dans la théorie de Holton.

Ils s'inscrivent du reste dans une conception beaucoup plus générale de la motivation, ce qui permet, entre autres choses, de mettre en perspective des idées philosophiques reçues comme l'idée que la faiblesse de volonté est un phénomène paradoxal ou énigmatique qu'il faut expliquer, plutôt que les mécanismes de contrôle de soi. Bien que cela ne soit pas une conséquence logique de la théorie du choix intertemporel, cette théorie oriente notre attention sur les capacités de contrôle de soi qu'acquièrent et raffinent les agents au cours de leur vie et qui leurs permettent en retour de repousser, voire de renoncer à, des gratifications importantes dans le futur.

Je me propose donc d'examiner dans ce chapitre les propriétés des choix intertemporels et ceux des décisions akratiques en particulier. Je tiens d'emblée à préciser que bien que les décisions akratiques exemplifient des propriétés temporelles spécifiques, ces propriétés forment des conditions nécessaires, mais non suffisantes pour diagnostiquer un épisode de faiblesse de volonté. Les décisions akratiques partagent leurs propriétés intertemporelles avec d'autres types de décisions. Ce qui les distingue toutefois se situe en amont au niveau de leurs motivations. Je discuterai de ces motivations dans le chapitre 10.

9.2 Les propriétés intertemporelles des décisions

Comme nous l'avons vu aux chapitres 2 à 4, les problèmes décisionnels présentent aux agents des alternatives de choix virtuellement praticables par lui. En décidant en faveur d'un choix, l'agent produit un certain nombre de conséquences pour lui-même et/ou pour les autres. Ces conséquences ne sont pas toujours ce qui importe exclusivement à l'agent dans sa sélection (ou élimination) de choix dans toutes les situations. S'il adhère à des principes moraux ou esthétiques pertinents au problème décisionnel, l'agent sélectionnera son choix en fonction des actions *per se* ou des propriétés intrinsèques qu'elles exemplifient et pas seulement en fonction de leurs conséquences. Bien que la réalisation de conséquences soit ce qui motive les agents à faire certains choix dans la plupart des situations, il semble qu'on n'en tient pas toujours compte dans l'*identification de leurs choix et de leurs décisions*. On dira plus naturellement qu'un agent a choisi une ligne de conduite, choisi d'emprunter un chemin, ou, de manière plus précise, choisi de faire ceci ou cela, tandis qu'il est plus artificiel de dire qu'un agent a choisi de produire un événement ou a choisi d'engendrer un effet spécifique.

Pourtant, une analyse de l'usage des termes d'action révèle qu'ils fournissent souvent une information sur des conséquences (Feinberg, 1970 ; Bratman, 1997) : tuer quelqu'un implique qu'on a causé sa mort, acheter un bien implique qu'on a causé sa vente, ouvrir une porte implique qu'on a causé l'ouverture de la porte, etc. Si l'on ne souhaite cependant pas mettre l'emphase sur certaines conséquences, on utilisera des termes d'action plus proches du vocabulaire descriptif des mouvements corporels. Au lieu de dire que Fulgence a choisi de tuer Réal, on dira, par exemple, qu'il a choisi de tirer sur la gâchette.

Pour ce qui est des choix ou décisions d'accomplir une action particulière, on a souvent besoin d'inclure des informations sur les conséquences pour les spécifier. Évidemment, lorsqu'on choisit d'accomplir une action spécifique, on ne choisit pas de les accomplir *ipso facto* sous toutes les descriptions qui mentionnent des informations sur ses conséquences. Choisir de tirer sur la gâchette n'implique pas qu'on choisisse de tuer quelqu'un en dépit du fait que tirer la gâchette a eu pour résultat que quelqu'un meurt. On peut décrire ce comportement comme un meurtre, mais on ne dira pas nécessairement qu'il relève du choix de tuer. Ce serait un choix de tuer seulement si la mort était un effet voulu, ou même seulement attendu, de l'agent. Le point à retenir cependant est que la description des choix

et des décisions peuvent – et doivent à l'occasion – inclure des considérations sur les conséquences. Il est dans certaines situations tout à fait approprié de les inclure et tout à fait inapproprié de ne pas le faire.

Ceci étant dit, une bonne description des choix akratiques devrait également inclure des considérations sur leurs conséquences. Aussi, à un niveau théorique, ces considérations doivent être suffisamment générales pour offrir une bonne conceptualisation de la faiblesse de volonté. Mais on doit faire attention de ne pas confondre les propriétés des conséquences des choix akratiques avec les motivations à les faire. Par exemple, dans la Théorie du choix rationnel, on fait abondamment mention des conséquences des actions, mais seulement comme conséquences *espérées* et en ne tenant compte que de leur probabilité *subjective* de réalisation. En œuvrant à l'intérieur de ce modèle, on demeure du côté de la motivation à faire des choix, et non du côté des conséquences effectives. Or, pour caractériser correctement les choix et les décisions, on doit également mentionner, quand cela est pertinent, des conséquences réelles qu'ils produisent. Dans le cas des décisions akratiques, la mention de ces conséquences est essentielle. Ce qui arrive à l'agent une fois qu'il a pris une décision akratique exemplifie des propriétés particulières qui entrent dans l'explication et la compréhension du phénomène.

Aussi, au niveau conceptuel, donner une place importante aux conséquences des choix akratiques nous permet de les envisager comme des *choix intertemporels*. Il s'agit d'une orientation conceptuelle que beaucoup de chercheurs en psychologie du choix et en économie comportementale ont prise ces dernières années. Les théories du choix intertemporelle offrent un éclairage nouveau sur le phénomène de la faiblesse de volonté. À l'intérieur de ce cadre conceptuel, beaucoup de chercheurs ont mis en relief des données expérimentales qui tendent à montrer que les agents qui vivent un épisode de faiblesse de volonté optent typiquement pour des allocations intertemporelles décrivant un schéma distinctif.

9.2.1 Les choix comme alternatives d'allocation intertemporelle intrasubjective

Qu'est-ce qu'un choix intertemporelle? *Grosso modo*, il s'agit d'un choix qui reflète en quelque sorte une solution de compromis entre les différentes récompenses et pénalités disponibles et subies à différents moments du temps (Loewenstein & Elster , 1992 ;

Frederick, Loewenstein & O'Donoghue, 2003). Plus concrètement, lorsqu'un agent fait face à des alternatives entre lesquelles il doit choisir, il évalue prospectivement – ne serait-ce que de manière sommaire – la *distribution temporelle des conséquences* découlant de chacune d'elle, et choisit entre ces distributions.

On peut conceptuellement traiter ces distributions comme des *allocations intertemporelles*. Quand on parle de problèmes d'allocations, on a souvent en tête des problèmes de distribution de ressources et de charges au sein d'une population. Disposant d'un sac de bonbons, je peux distribuer ceux-ci à un groupe d'enfants suivant une règle d'allocation particulière. Cette règle peut être plus ou moins complexe, mais elle spécifiera qui doit avoir des bonbons, combien de bonbons ils pourront avoir, et quelle sorte – dans l'éventualité où certains sont meilleurs que d'autres. Elle devra correspondre à la préférence de certains enfants et non pas d'autres, etc. Mais je peux également, dans un geste égoïste, garder les friandises pour moi seul. Dans ce cas, toute décision de consommation sera une allocation intertemporelle dans la mesure où je ne serai pas indifférent aux instants de ma consommation. J'étendrai alors ma consommation dans le temps en périodes rapprochées ou distantes. Je choisirai combien de bonbons manger en séries ininterrompues, lesquels manger avant, lesquels manger après, en fonction notamment de l'intensité du plaisir qu'ils me procurent si je les ai consommés dans tel ou tel ordre. Mais je peux aussi décider de manger une série tout de suite et me garder les autres pour une période de consommation indéterminée.

Ce qui est important de comprendre ici est que les allocations intertemporelles qui sont intéressantes pour une conception de la faiblesse de volonté sont des allocations dont le principal tributaire est l'agent qui en est lui-même responsable, bien que certaines des conséquences affecteront plutôt d'autres organismes ou des objets inanimés. Par ricochet, une décision d'allocation intertemporelle peut affecter indirectement l'agent qui l'a prise, comme dans le cas où je fais une promesse à quelqu'un. La promesse crée une attente chez autrui qui le motivera ultérieurement à demander des comptes ou à me punir en cas de défection. Aussi, faire une promesse modifie le caractère désirable de certaines lignes de conduite qui seront disponibles dans l'avenir.

On peut isoler des caractéristiques générales et pertinentes des allocations de la manière suivante :

1/ **Le profil expérientiel :** La classe des conséquences qui importe le plus est celle des conséquences qui affectent directement ou indirectement l'agent qui procède à l'allocation. Ariely et Carmon (2000) appellent *profil expérientiel* la manière dont ces conséquences distribuées dans le temps affectent l'agent. Un profil expérientiel n'inclut pas nécessairement l'ensemble des conséquences qui affectent effectivement la structure de l'organisme de l'agent, mais plutôt celles qui ont un impact hédonique suffisant pour qu'elles puissent (en principe) apparaître sur son radar mental. Boire une tout petite gorgée de jus de fruit a une conséquence hédonique évidente à très court terme, mais engendrera aussi un apport calorique minimal ultérieurement, et dont l'impact hédonique sera peut-être nul. Le meilleur moyen de mesurer le profil expérientiel d'une allocation reste le rapport verbal que les agents sont disposés à faire pour chaque instant de la période où ils « subissent » la distribution. Mais les tests psychométriques plus sophistiqués peuvent être nécessaires pour détecter, dans certains cas, la présence d'émotions inconscientes comme Berridge et Winkielman (2004) l'ont montré. Dans tous les cas, le profil expérientiel d'une allocation est établi en fonction de plusieurs paramètres et comporte habituellement des informations sur le moment et la durée d'une expérience hédonique, sa qualité (plaisante, déplaisante ou neutre) ainsi que son intensité. Je ne dis pas ici que tout ce qui importe lorsque vient le temps de prendre une décision d'allocation intertemporelle réside dans son profil expérientiel, mais seulement que ce sont les considérations hédoniques qui sont pertinentes dans la caractérisation de la faiblesse de volonté.

2/ **La possibilité et le coût de révision :** Si l'on accepte la description que je viens de donner des allocations intertemporelles, il va sans dire qu'elles ne sont pas des décisions inéluctables et elles peuvent dans certaines situations être modifiées *a posteriori*. Si je prends une décision d'allocation celle-ci engendrera causalement des conséquences pour moi-même. Certaines de ces conséquences sont inéluctables, alors que d'autres peuvent être limitées ou accentuées par ma propre intervention ou par une intervention extérieure. Contracter une hypothèque pour l'achat d'une propriété immobilière engendre dans le monde extérieur des attentes de paiements légalement contraignantes et des risques de saisies ou de dévaluation d'une cote de crédit. Une fois que la transaction et l'hypothèque sont officiellement cautionnées par les parties dans le cabinet d'un notaire, ces conséquences sont automatiquement engendrées si personne ne fait rien pour qu'il en soit

autrement. Sans pouvoir revenir comme tel sur ma décision, je peux néanmoins, sous certaines conditions, revoir la distribution des charges et m'entendre avec mon institution financière concernant les modalités de paiement. Je peux à la rigueur revendre la propriété si je juge que ces paiements représentent finalement un fardeau trop grand. Mais il y a toujours un coût pour parer aux conséquences d'une allocation intertemporelle. En ce sens, une fois qu'on a pris une décision, il faut vivre avec les conséquences. Si j'ai choisi des modalités de paiement en achetant une maison en pensant que je pourrai ultérieurement les changer, je dois accepter le fait que je devrai dépenser un peu de temps et d'énergie pour rencontrer le représentant de l'institution financière et le convaincre de bien vouloir modifier mes modalités de paiement. Mais, si je veux que les attentes légales de paiement, que ma décision a engendrées, cessent tout à fait, je devrai soit vendre ma maison, soit déclarer faillite. Dans les deux cas, le coût de modification est énorme. Or, la possibilité de modification d'une allocation et son coût subséquent est un aspect particulièrement important pour comprendre les cas de faiblesse de volonté qui relèvent d'un changement d'avis irrationnel ou de la non-observance d'une résolution. Mais il s'agit aussi d'un aspect important pour expliquer comment et pourquoi les agents sont capables de maintenir le cap dans un projet, de résister aux tentations et de respecter les délais, notamment au moyen d'engagements préalables.

3/ **Le caractère plus ou moins vague :** Cela dit, la distribution des conséquences ne doit pas toujours être vue comme quelque chose de précis ou de bien défini. Les conséquences distribuées s'avèrent souvent diffuses. Contracter une hypothèque engendre une ponction monétaire récurrente, à une date précise, voire à une heure précise. Prendre une bière de trop provoque des haut-le-cœur lancinants et assure un mal de tête pour le lendemain. Cependant, les conséquences d'une allocation peuvent ne pas être en « acte », mais seulement en « puissance ». C'est le cas du potentiel de risque que représente la consommation de cigarettes et le potentiel de jouissance que comporte la possession d'un bien ou de moyens financiers, même s'ils ne seront jamais utilisés ou dépensés.

4/ **L'impact résiduel :** Les décisions d'allocation engagent souvent les agents dans des activités qui ont un impact hédonique positif immédiat, mais presque aucun impact résiduel. Manger une pomme, se gratter l'oreille, écouter une pièce de musique ou contempler un coucher de soleil sont des activités qui peuvent être voulues et appréciées au

moment où elles sont accomplies, mais elles ne laissent que rarement des traces hédoniques résiduelles. Elles n'apparaissent habituellement pas comme un chapitre de l'histoire d'une personne. Par contraste, il y a des activités qui sont voulues par les agents en vertu de leurs conséquences hédoniques, mais dont l'impact positif est purement résiduel. Se soumettre à une intervention dentaire, lutter pour obtenir une promotion, faire une demande de subvention, payer son compte de taxes, nettoyer son garage sont des activités qui n'ont un impact hédonique positif qu'une fois accomplies, et souvent un impact négatif pendant qu'elles sont accomplies. Étant donnée l'ampleur de l'impact hédonique résiduel, certaines de ces activités auront une signification particulière pour l'agent et occuperont par le fait même une place importante dans son histoire.

5/ **Le caractère non nécessairement conscient :** conçues de cette manière, on se rend compte peu de décisions ne relèvent pas d'une forme ou d'une autre d'allocation intertemporelle. Les cas paradigmatiques d'allocations intertemporelles sont se marier, choisir un emploi, poursuivre ou non des études, entamer un régime minceur, placer de l'argent, procréer ou fumer. Pour peu que les agents fassent face à des choix dont les conséquences pertinentes sont distribuées dans le temps ou perdurent dans le temps, leur décision relèvera d'un choix intertemporel. D'ailleurs, cela ne nous empêche pas de considérer que des organismes qui n'ont pas une conscience du déroulement du temps effectuent néanmoins ce genre de choix. La condition étant que la valeur de ces choix, mesurée en facteur de *fitness* – la capacité d'engendrer un maximum de copies de son génome par le biais de la reproduction –, dépende de variables dont la valeur sera déterminée dans le futur. Dans un article sur les facteurs évolutionnels qui ont mené à la formation de mécanismes de patience, Kaselnik (2002) suggère que même des êtres qui sont complètement myopes à l'égard du futur font également des « choix » intertemporels. Au printemps, un arbre doit déterminer combien de matière ligneuse et de nouvelles branches il doit amorcer la pousse, mais aussi combien de feuilles et de fleurs. Le degré de pollinisation par les insectes, la météo, la quantité d'oiseaux disséminant des graines, le croisement des arbres adjacents sont autant de facteurs cruciaux pour déterminer la ou les bonnes allocations. Mais, contrairement aux agents conscients, ces choix sont entièrement déterminés par des mécanismes aveugles implémentés par la sélection naturelle. D'ailleurs, seuls les agents disposant de circuits neuronaux pouvant produire du plaisir et du déplaisir

sont susceptibles de choisir des distributions en fonction de considérations hédoniques. Mais la présence de ces circuits dans un organisme ne s'accompagne pas nécessairement de structures neuronales responsables d'une forme, même primitive, de conscience de soi. Si la faiblesse de volonté consiste en un schéma d'allocation intertemporelle inadéquat, cela autorise en théorie d'observer des cas de faiblesse de volonté animale qui semblent, à première vue, assez loin des cas humains.

6/ **Les formes pertinentes des allocations** : Bien qu'il n'y ait pas, à ma connaissance, de typologies détaillées des allocations dans la littérature psychologique et économique, on peut estimer qu'il y a 3 types fondamentaux : (1) il y a les allocations qui concentrent les conséquences hédoniquement positives dans le futur immédiat et les conséquences négatives dans le lointain ; (2) les allocations qui distribuent plus uniformément les conséquences ; (3) puis, les allocations qui concentrent leurs conséquences négatives dans le futur immédiat et les conséquences positives dans le lointain. De manière plus simple, on dira que l'allocation du premier type présente un profil hédonique *décroissant*, celle du second type, un profil *uniforme*, et celle du troisième type, un profil *croissant*. Les schémas d'allocation en faveur desquels optent les agents akratiques correspondent à des types d'allocation avec profil hédonique décroissant.

9.2.2 Les schémas d'allocation typiques de la faiblesse de volonté

Pour déterminer si l'on a affaire à un épisode de faiblesse de volonté, on doit d'abord examiner les allocations intertemporelles *disponibles* pour l'agent. Si parmi toutes les allocations disponibles pour lui, il en a choisi une qui est clairement décroissante, on peut suspecter un épisode de faiblesse de volonté. En fait, les épisodes d'akrasie décrivent des schémas d'allocation typiques, assez simples, et qu'on peut aisément reconnaître. Dans chaque cas, on observe que l'agent fait passer ses intérêts à court terme devant ses intérêts à long terme, et cela, en dépit du fait que ces derniers pèsent beaucoup plus lourd dans le profil hédonique de l'agent que les premiers. Aussi, l'agent akratique choisira une allocation qui présente un profil hédonique globalement inadéquat. Plus précisément, une décision d'allocation typique de faiblesse de volonté présentera des conséquences positives supérieures en quantité et en intensité dans un avenir rapproché, mais des conséquences négatives supérieures en quantité et en intensité dans un avenir plus distant. C'est, pour

ainsi dire, l'inverse d'une décision d'investissement. L'agent akratique choisira cette allocation inadéquate en dépit du fait qu'il y a une autre allocation intertemporelle disponible pour lui qui présente un profil hédonique supérieur. Les allocations supérieures, négligées par l'agent akratique, peuvent présenter des profils hédoniques plus uniformes ou des profils qui impliquent des récompenses beaucoup plus importantes, mais susceptibles de l'affecter lors d'une période temporellement plus distante. C'est donc en vertu du fait que certaines décisions peuvent être conçues comme des décisions d'allocation intertemporelle et que ces décisions présentent des schémas de distribution particuliers qu'on peut les diagnostiquer comme des épisodes de faiblesse de volonté.

Par exemple, si je crois avoir une carie qui produit un mal récurrent, je considérerai alternativement l'option d'aller chez le dentiste et l'option de m'en abstenir. Je peux me résoudre à me soumettre à une intervention dentaire douloureuse et à prendre un rendez-vous plus tard dans la semaine avec mon dentiste. Le temps passe et, une heure avant de subir l'intervention, je reconsidère ma décision et évalue de nouveau les options en lice : soit me soumettre à l'intervention ou m'en abstenir. Chacune de ces options engendre une distribution particulière d'épisodes plaisants et déplaisants, et même hédoniquement neutres. Décider de se soumettre à l'intervention produira une période intensément déplaisante lors de l'obturation de ma dent et modérément déplaisante pour le restant de la journée, cela étant dû à l'irritation de mes terminaisons nerveuses. Le lendemain sera parsemé de beaucoup de moments d'inconfort lorsque mon attention se portera sur ma condition dentaire. Puis ces moments d'inconfort seront de plus en plus rares et finiront par disparaître complètement dans une semaine. Si j'opte pour une abstention plus ou moins momentanée, je m'expose à ressentir un mal lancinant plusieurs fois dans les journées à venir et qui sera de plus en plus intense et de plus en plus fréquent. Mais en m'abstenant, j'obtiendrai un soulagement momentané. Ce soulagement sera d'autant plus intense que j'ai angoissé une bonne partie de la semaine du fait que j'anticipais en imagination un traitement de canal pour une molaire et que, par effet de contraste, j'améliorais du tout au tout ma condition affective à court terme.

Dans cette situation, j'ai le choix entre deux allocations intertemporelles qui ont des structures complètements différentes. Si j'opte pour l'abstention, je choisis alors une allocation qui concentre la plupart des conséquences positives dans les instants suivant ma

décision. Mais les conséquences négatives s'étaleront sur une période beaucoup plus longue et iront sans doute en s'accentuant. Mais j'éviterai aussi des conséquences hautement négatives à court terme découlant directement d'une intervention dentaire pénible, en dépit du fait que leurs impacts résiduels sont minimes – on est rarement traumatisé par une visite chez le dentiste.

Les deux allocations imaginées décrivent des schémas typiques de choix binaire qui comportent un risque important pour l'agent qui souhaite améliorer son sort à long terme. On peut représenter ces schémas compétitifs en les simplifiant à l'extrême à l'aide du diagramme suivant :

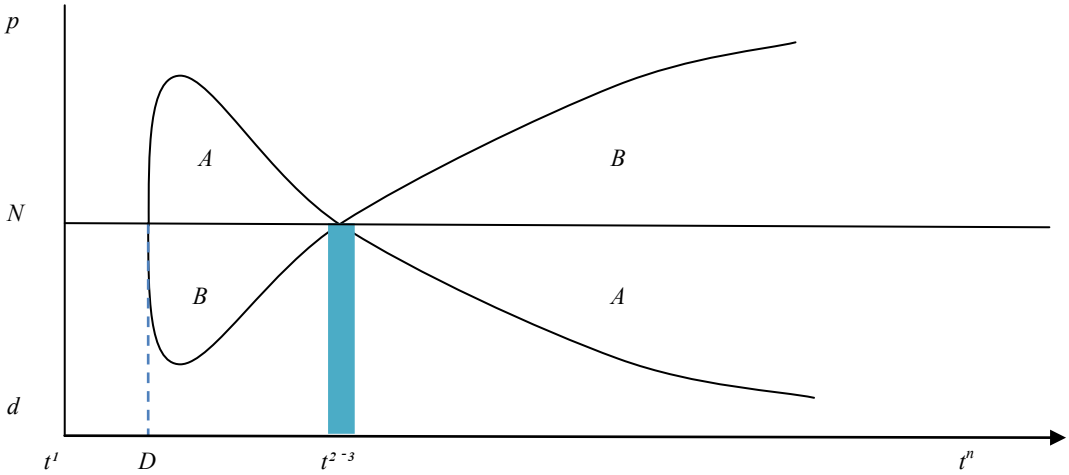

Figure 2

Ce diagramme décrit le profil hédonique de deux allocations intertemporelles *A* et *B* disponibles en même temps sous forme de choix binaire. Le diagramme décrit deux profils établis sur la durée, en fonction d'un point de référence hédonique neutre *N* : au dessus de *N* les conséquences sont plaisantes, en dessous, elles sont déplaisantes. *D* représente un point de décision et t^2 l'horizon temporel du court terme. Par souci de simplicité, les deux allocations exemplifient des moments de plaisir et des moments de déplaisir contigus. Les cas réels présentent des profils expérientiels bigarrés où la portion temporelle correspondant au court terme est parsemée d'instants de plaisir et de déplaisir.

Le diagramme montre que si l'on décide de s'abstenir d'aller chez le dentiste, alors qu'on a manifestement un problème dentaire important, on opte pour une allocation dont les conséquences positives se concentrent dans la portion correspondant au court terme, tandis que les conséquences négatives seront éprouvées sous forme résiduelle sur une période

subséquente. Le plaisir découlant d'une abstention serait initialement intense – il peut commencer au moment où l'on pensera prendre une décision et non seulement au moment où l'on prend effectivement la décision – et décroîtra jusqu'à atteindre la frontière de l'état homéostatique neutre, puis déchoira dans un déplaisir croissant.

Ici, l'horizon du court terme est évidemment variable et déterminé – souvent vaguement – par la nature et l'étendue des allocations, ainsi que les allocations (ou plans d'ensemble) plus globales dans lesquelles elles s'inscrivent. Le court terme à l'intérieur d'une journée n'est pas le court terme à l'intérieur d'une année. Pour l'exemple qui nous intéresse, l'horizon du court terme s'étendra sur quelques heures. Aussi, la frontière du court terme ne se situe pas toujours à l'instant où la valeur hédonique des conséquences change de qualité – ici en t^2. Mais c'est souvent à cet instant que les agents éprouvent le regret de ne pas avoir choisi l'allocation supérieure et considèrent alors avoir décidé sur la base d'une surévaluation des considérations à court terme.

Le problème pour l'agent qui choisit l'allocation *A* dans ce genre de situation n'est pas tant que la quantité et/ou l'intensité des instants plaisants est globalement inférieure à celles des instants déplaisants. Un agent peut ne faire face dans certaines situations qu'à des allocations globalement désavantageuses. Aussi, le problème résidera dans le fait qu'une autre allocation disponible présente un profil hédonique global supérieur[101].

[101] Bien que cela soit assez intuitif, on obtiendra, de manière rigoureuse, la valeur d'un profil hédonique en multipliant l'intensité du plaisir par sa durée, et l'on soustraira du résultat le produit de l'intensité du déplaisir et de sa durée. Établir un profil hédonique complet n'est évidemment possible qu'en utilisant des mesures psychométriques répétées à intervalles rapprochés, et cela, dans le contexte d'une intervention très limitée dans le temps. C'est ce qu'ont fait Ariely et Carmon (2003) lorsqu'ils ont voulu mesurer l'impact hédonique d'interventions médicales douloureuses comme la colonoscopie. À intervalles rapprochés, les patients qui subissaient l'intervention devaient rapporter l'intensité de leur état de douleur sur une échelle de 0 à 100, où 0 représente un état neutre ou sans douleur, et 100, la pire douleur qu'ils puissent imaginer. Les chercheurs pouvaient obtenir un profil hédonique précis. Cependant, lorsqu'ils ont demandé au patient de faire un bilan hédonique de son intervention une fois terminée, ceux-ci établissaient leur estimation d'une manière complètement différente et, par conséquent, arrivaient à des résultats souvent différents. Contrairement à ce qu'on aurait tendance à faire si l'on disposait d'une vision rétrospective détaillée de nos moments hédoniquement chargés, ils se basaient essentiellement sur 3 variables pour déterminer le niveau de douleur globale : l'intensité du pic de douleur, l'intensité de la phase finale, ainsi que la pente globale (ascendante ou descendante) de l'expérience totale. Cette heuristique d'évaluation rétrospective du profil hédonique est utile dans la mesure où les agents ont des capacités mnésiques limitées, mais doivent tout de même faire des bilans hédoniques pour s'orienter dans l'avenir. Mais cette heuristique produit ce que les chercheurs nomment le biais de négligence de la durée. L'impact d'interventions médicales beaucoup plus longues, mais qui présentent un pic de douleur et une finale moins intense, ainsi qu'une pente descendante sera systématiquement minimisé bien qu'il puisse être globalement supérieur. En outre, ceci montre bien qu'on ne doit pas seulement distinguer les profils hédoniques des attitudes prospectives que les agents ont à leur égard, mais également des attitudes rétrospectives.

Dans beaucoup de situations, les agents doivent décider entre des allocations qui concentrent tous les effets bénéfiques sur le court terme, mais à intensité différente. Un agent akratique optera pour celles qui présentent un profil hédonique supérieur sur l'horizon du court terme en dépit du fait qu'une autre allocation comporte beaucoup moins d'inconvénients à long terme et que ces inconvénients pèsent plus lourd dans la balance que l'ensemble des bénéfices. Il n'est pas irrationnel en soi de choisir une option qui concentre ses bénéfices sur le court terme et relègue les inconvénients à plus tard. S'abonner à un service de câblodistribution, contracter un prêt automobile ou une hypothèque, utiliser sa carte de crédit ou opter pour une formule du type « Achetez maintenant, payez plus tard! », peuvent être à l'occasion des choix judicieux. Ceci s'explique par le fait que bon nombre de biens et de services ne sont disponibles que sous ces modalités. D'ailleurs, comme les revenus et les salaires des personnes ne sont pas versés sur une base journalière, mais le plus souvent hebdomadaires ou bimensuels, cela pose certaines contraintes temporelles que les formules de financement et d'abonnement exploitent. Mais ce n'est évidemment pas parce qu'il peut être judicieux de choisir parmi des options qui exemplifient une telle structure d'allocation qu'un agent est à l'abri d'un choix akratique. Il y a une différence importante, du point de vue du choix rationnel, entre faire, à crédit, une épicerie frugale et faire, à crédit, une épicerie gloutonne lorsque nos sources de revenu futur sont plutôt maigres. Un agent akratique optera typiquement pour un usage peu modéré des diverses formules de financement même dans les situations où il n'a pas vraiment le choix d'en utiliser une. Il optera pour une formule qui représente un profil hédonique supérieur sur l'horizon du court terme et négligera l'option qui engendre moins d'inconvénients au-delà de cet horizon, en dépit du fait que cette dernière présente un profil global plus avantageux (figure 3). Un agent akratique choisira, par exemple, un forfait de câblodistribution coûteux qui offre une quantité trop importante de chaînes pour la consommation qu'il en fait, choisira des options coûteuses pour une automobile neuve, achètera une propriété immobilière trop grosse et dispendieuse, parce que les coûts souvent prohibitifs de ces choix ne seront assumés que dans un avenir plus ou moins lointain – à la fin du mois ou de l'année, ou même après plusieurs années[102].

[102] Ce qu'on a appelé la crise des « *sub-prime* » aux États-Unis en 2008 fut en partie causée par des achats de propriétés à des prix exorbitants et des produits hypothécaires à taux croissants ou à taux nuls pour les premières années.

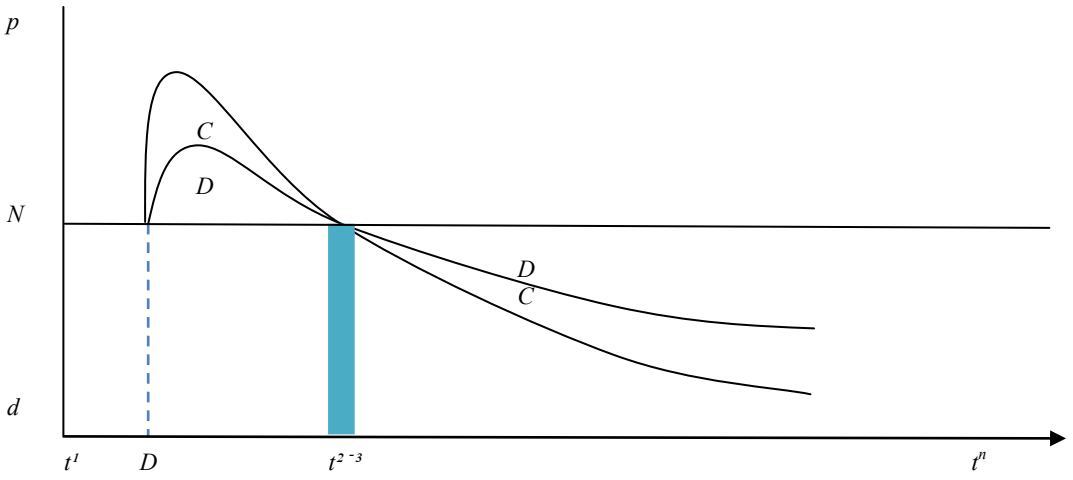

Figure 3

Les investissements présentent des structures d'allocation différentes. Il y a évidemment toutes sortes de formules, et il est souvent difficile de s'y retrouver. Mais en général, les investissements impliquent de payer maintenant et de recevoir plus tard. Aussi, un agent peut exhiber un problème de faiblesse de volonté même en prenant une décision d'investissement. Si une femme fait pression sur son mari pour qu'il investisse une partie de leurs revenus dans un fond de retraite, celui-ci peut choisir un plan mensuel qui n'implique qu'une somme minimale, et qui, par conséquent, leur offrira moins de revenus pour leurs vieux jours. Mais il n'est pas nécessaire d'être contraint à faire un investissement pour qu'une allocation de ce type soit engendrée par une décision akratique. Décider d'investir une partie de ses revenus au lieu des les dépenser maintenant n'est pas un choix akratique. Décider d'investir dans un plan qui implique le minimum en coût de renoncement actuel au lieu d'un plan plus avantageux sur le long terme est probablement akratique. Dans ce cas, l'agent akratique choisit une allocation qui présente le profil hédonique le moins désavantageux sur l'horizon du court terme (figure 4).

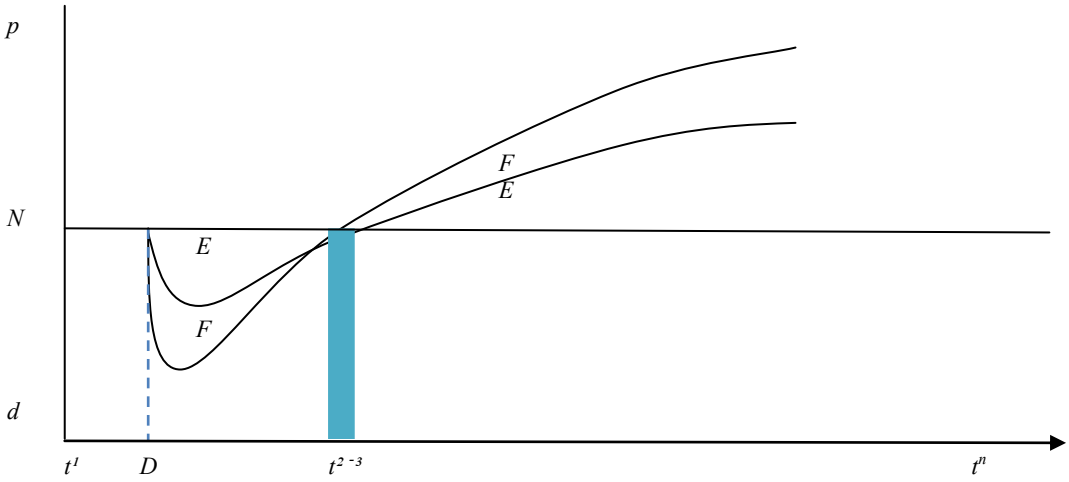

Figure 4

La procrastination structurée, dont le philosophe John Perry fait l'éloge sur son site humoristique, décrit le même profil hédonique que les cas de sous-investissements[103]. Une procrastination peut être structurée lorsque les agents qui remettent une tâche ennuyeuse à plus tard s'acquittent néanmoins, et de manière systématique, de besognes un peu moins ennuyeuses. La procrastination ne s'accompagne pas nécessairement de comportements impatients. Lorsqu'on a des formulaires administratifs à remplir, un tuyau qui fuit à colmater, une pile de chefs-d'œuvre de ses étudiants à corriger ou une déclaration d'impôt sur le revenu à compléter, on est, par magie, beaucoup plus disposé à faire du rangement, passer l'aspirateur, nettoyer le four de la cuisine ou astiquer son instrument de musique. Le plaidoyer que fait Perry de la procrastination structurée repose sur l'idée qu'il s'agit de la meilleure façon d'utiliser, à notre compte, nos propres défauts en matière de rationalité pratique. D'ailleurs, pour Perry, la procrastination structurée est rationnellement justifiable parce qu'elle relève d'une forme de perfectionnisme (Steel, 2007). Les personnes perfectionnistes sont souvent démotivées par l'ampleur des tâches qu'ils se donnent au quotidien et dans leur vie professionnelle. Les objectifs trop ambitieux engendrent souvent une sérieuse inhibition. La réponse à cette inhibition se traduit souvent par une autorisation

[103] Structuredprocrastination.com : 1995-2008.

pour soi-même à effectuer de manière imparfaite une autre tâche qui occupe une place plus périphérique dans notre plan.

9.2.2.1 Autres schémas d'allocation moins courants

Tous les problèmes de faiblesse de volonté ne présentent pas nécessairement des profils hédoniques potentiels comme ceux décrits par les figures 8 à 10. Il faut tenir compte des situations où l'agent doit décider entre des allocations qui ne comportent que des bénéfices ou que des coûts. Dans beaucoup d'expériences de laboratoire, on propose à des agents humains ou animaux de choisir entre deux biens de valeur différente, mais disponibles à des instants différents, et non pas entre des biens qui se transforment en maux et des maux qui se transforment en biens. Dans une expérience célèbre, Rachlin et Green (1972) offrent à des pigeons la possibilité d'appuyer sur des touches qui actionnent un mécanisme qui distribue des quantités de nourritures variables à des moments différents. Thaler a mesuré, pour la première fois, les effets qu'avait sur des humains l'offre de deux récompenses différentes en argent, disponible à des instants différents (Thaler, 1981). Je ne connais toutefois pas d'expériences qui mettent les agents face à des dilemmes (de nature morale ou non) qui ne comportent que des conséquences négatives. Mais on peut raisonnablement croire que les agents choisissent de la même manière que dans les situations où ils ne font face qu'à des options avantageuses, et qu'on peut également y voir des épisodes de faiblesse de volonté.

Je ne sais pas toutefois si les contextes de choix qui ne présentent que des options avantageuses ou désavantageuses sont répandus. Beaucoup de problèmes décisionnels concoctés en laboratoire mettent les agents face à un dilemme de « gains contre gains ». Au-delà du fait que ces situations sont artificielles et ne correspondent pas à des problèmes décisionnels qu'on rencontre à l'extérieur des laboratoires, on peut émettre des réserves sur la description adéquate de ces allocations « gain contre gain » ou « perte contre perte ». On a l'intuition que de renoncer à un gain plus petit, mais rapproché dans le temps, à la faveur d'un gain plus grand, mais plus distant, comporte sa part de désagréments. Si je vous offre 20 $ maintenant ou 25 $ dans une semaine, vous pourriez ressentir un léger désagrément à l'idée de devoir renoncer à un gain immédiat. Dans ce cas, l'allocation aura un profil hédonique ressemblant au profil *B* (figure 2). Si, en plus, on peut s'attendre à ce que le choix d'allocation *C* produise un décroissement de la valeur hédonique positive, qui déchoit

en déplaisir en raison du regret subséquent de ne pas avoir choisi le gain plus grand, alors le profil hédonique correspondra comme tel à *A* (figure 2). C'est pour ces raisons qu'on peut, je pense, considérer que les choix d'allocations typiques auxquels fait face l'agent akratique présentent des profils hédoniques schématisés dans les figures 2 à 4. Dans quelle mesure il existe des contextes de choix temporels qui présentent des alternatives de purs gains ou de pures pertes est une question empirique pour laquelle il est difficile de fournir une réponse. Le simple fait que les agents puissent se remémorer des choix passés et ressentir des regrets à leur égard brouille certainement les cartes.

Que cela existe et soit répandu ou non, l'important se situe dans le fait qu'un agent qui vit un épisode de faiblesse de volonté opte pour une allocation qui concentre ses bénéfices à l'intérieur d'un horizon de court terme et délaie la majeure partie de ces coûts au-delà de cet horizon.

Ceci, je pense, permet d'expliquer les deux catégories fondamentales de la faiblesse de volonté, à savoir l'impatience et la procrastination. On est impatient lorsqu'on décide de faire quelque chose qu'on aurait pourtant avantage à reporter dans le temps. La procrastination est, à première vue, un phénomène inverse. Elle consiste à rapporter dans le temps quelque chose qu'on pourrait accomplir avant. Mais on aurait tort de penser que la procrastination est rigoureusement le phénomène inverse de l'impatience. Lorsqu'un agent procrastine ou se montre impatient, il opte en fait pour une allocation intertemporelle qui a la même structure. Certes, les activités qu'on est disposé à remettre à plus tard ne sont pas les activités qu'on est impatient d'accomplir. On procrastine lorsque vient le temps de faire des besognes ou de s'acquitter de tâches ennuyeuses, et l'on s'impatiente lorsque vient le temps de s'adonner à des loisirs ou à des activités hautement excitantes et grisantes. Mais même si les besognes et les loisirs présentent des structures d'allocation différentes, l'allocation intertemporelle a la même structure. Habituellement, si l'on s'acquitte de certaines besognes, ce n'est pas pour des raisons masochistes, mais bien pour leur impact résiduel positif : *on n'aime pas faire la vaisselle, mais on aime que la vaisselle soit faite*. De même, on ne vide pas son compte de banque en achetant des biens pour des raisons masochistes : *on n'aime pas que l'argent soit dépensé, mais on aime dépenser de l'argent*. Aussi, dans ces deux cas, la partie de l'allocation qui importe le plus pour nous – en tant

qu'agents procrastinateurs et impatients – est l'ensemble des conséquences *cooccurrentes* à l'activité et non l'ensemble des conséquences qui *suivent* l'activité.

La structure des allocations intertemporelles ne permet pas seulement de fournir les bases d'une analyse conceptuelle de la faiblesse de volonté, elle permet aussi de faire un classement raisonné des activités susceptibles de présenter un risque pour l'agent à la volonté faible. Les théories et conceptions qu'on trouve dans la littérature philosophique négligent l'examen des caractéristiques temporelles des alternatives représentant un problème décisionnel. Aussi, on ne tente pas d'expliquer pourquoi on ne peut pas devenir dépendant à une substance qui produit des effets négatifs aigus comme des étourdissements, des nausées, des sueurs froides, des courbatures et une migraine dans une période qui suit immédiatement sa métabolisation, mais qui produit des effets agréables, comme la sensation de flotter, l'ivresse, l'absence de soucis et d'inhibition et l'euphorie, beaucoup plus tard[104]. Et pourtant, on peut devenir tout à fait dépendant à des substances qui ont les mêmes effets, mais dont la distribution temporelle est inversée. L'explication intertemporelle peut fournir l'explication. Choisir entre deux substances qui ont les mêmes effets positifs et négatifs, mais distribués dans un ordre temporel différent, revient à choisir entre deux allocations complètement différentes.

Cela dit, la structure de l'allocation des drogues est en général telle qu'on ne procrastine pas lorsque vient le temps de s'en administrer une dose, qu'on soit un toxicomane ou un patient qui vient de subir une opération qui contrôle, dans une certaine mesure, son administration – comme c'est maintenant le cas dans bon nombre d'hôpitaux où on laisse les patients s'administrer des doses de morphine. En revanche, on procrastine lorsque vient le temps de s'administrer des antibiotiques mauvais au goût à des moments où l'on se sent bien et dont les conséquences bénéfiques ne sont pas aussi importantes au début du traitement. Les approches conceptuelles qu'on trouve dans la littérature philosophique se concentrent exclusivement sur le fonctionnement et les caractéristiques des processus

[104] D'ailleurs, une des hypothèses les plus probantes pour expliquer pourquoi la méthadone – un dérivé synthétique de l'opium qui est utilisé pour sevrer les héroïnomanes – n'est pas une substance aussi addictive que l'héroïne malgré le fait que ses effets hédoniques sont sensiblement les mêmes est que son temps de métabolisation est plus long et produit donc des effets plaisants plus tard et de manière plus diffuse. En fait, le temps de métabolisation des drogues est en partie déterminé par le mode d'absorption. L'ingestion produit une métabolisation moins rapide et engendre un profil hédonique plus diffus que l'inhalation, et cette dernière produit une métabolisation plus lente et engendre un profil plus diffus que l'injection par intraveineuse (Ainslie, 2001).

cognitifs internes de l'agent et négligent l'examen des caractéristiques des choix et la structure des problèmes décisionnels. Les processus décisionnels des organismes reflètent pourtant la structure des problèmes auxquels ils font face et ils sont d'autant plus efficaces quand ils y sont bien adaptés[105].

9.2.2.2 Le problème des mauvaises habitudes et des décisions à impact négligeable

Lorsqu'on traite des conséquences des choix akratiques on a souvent à l'esprit les impacts négatifs (réels ou probabilistes) très marqués de décisions isolées. Décider de faire sa déclaration de revenus après la date butoir, prendre le volant en état d'ivresse, boire de l'eau de mer lorsqu'on est assoiffé, annuler un rendez-vous chez le dentiste alors qu'on a mal aux dents, comportent des désavantages évidents. Mais que dire de la décision de fumer *une* cigarette ou de se régaler d'*une* part de gâteau?

En fait, si on met dans la balance les alternatives de fumer une seule cigarette et contrôler son envie de le faire, ou manger une seule part de gâteau et accomplir un acte d'autocontrôle diététique, il appert que les premières sont des alternatives qui produisent un meilleur profil expérientiel. Aucune décision de fumer une seule cigarette n'accroît significativement les risques pour la santé, de même qu'aucune décision de manger une part de gâteau n'altère la silhouette. Comme je l'ai mentionné au chapitre 6, rien ne permet d'exclure *a priori* que des décisions akratiques puissent relever d'une maximisation du bien-être en bonne et due forme. Le profil hédonique d'une décision akratique peut être optimal si celle-ci a peu ou pas d'impact résiduel.

Comme Elster l'a déjà fait remarquer, à cet égard certaines décisions akratiques posent un problème analogue pour toute théorie de la rationalité pratique au fameux paradoxe du vote (Elster, 1989). Il est difficile d'expliquer pourquoi les citoyens pris individuellement se déplacent pour voter alors qu'ils savent que leur vote ne changera pas l'issue du scrutin. Du moins, il est difficile de l'expliquer dans un cadre conséquentialiste. Beaucoup de théoriciens ont voulu relever le défi que pose ce paradoxe pour le conséquentialisme. Ce n'est pas le lieu pour discuter de ces réponses. Un auteur comme Boudon les a justement critiquées et en est venu à la conclusion que le paradoxe du vote démontre clairement que le

[105] Point que j'ai évoqué et exposé un peu plus en détail au chapitre 7.

conséquentialisme est erroné (Boudon, 1999, 2003). Est-ce que cela dément notre proposition selon laquelle les décisions akratiques présentent typiquement un profil expérientiel décroissant ou potentiellement décroissant?

La distribution des conséquences et leur profil hédonique demeurent une caractéristique centrale des épisodes de faiblesse de volonté. Seulement, dans certains cas, elles ne sont pas engendrées par des décisions isolées, mais par des groupes ou des séries de décisions d'un même type. Ce qu'on appelle communément les mauvaises habitudes présente des profils hédoniques désavantageux, sans quoi on ne les jugerait pas mauvaises. Aussi, si l'on veut déterminer si une décision ou un choix relève d'un épisode de faiblesse de volonté, on doit à l'occasion ne pas regarder seulement du côté de son impact résiduel. On doit également déterminer si la décision s'inscrit dans une « politique » ou une habitude qui, elle, a un impact résiduel. Par exemple, fumer la cigarette ou consommer de la pornographie ne relève pas comme tel de la faiblesse de volonté. Initialement, on commence à fumer pour faire de nouvelles expériences, pour s'intégrer à un groupe, pour défier l'autorité parentale ou pour d'autres raisons qui n'ont rien à voir avec le caractère addictif de la nicotine. Il ne s'agit pas de faiblesse de volonté. Ça le devient par contre au moment où fumer devient une habitude. De la même façon, consommer de la pornographie à l'occasion ne relève pas d'épisodes de faiblesse de la volonté. C'est seulement quand cela devient une habitude qui contamine les autres aspects de la vie qu'on entre dans l'akrasie.

Du reste, le problème des mauvaises habitudes et des dépendances est particulièrement intéressant pour une théorie de l'allocation intertemporelle. Les dépendances à des substances (drogues, alcool, café, etc.) ou à des activités (jeux, pornographie, télévision, travail, etc.) ainsi que toute une panoplie de mauvaises habitudes produiront, à plus ou moins long terme, une accentuation de la décroissance de leur profil hédonique. Le processus de « *sensibilisation* » du système dopaminergique, tel que décrit par Berridge et Robinson (1993, 2003, 2008), ne rend pas seulement certains stimuli plus saillants pour la motivation. Il s'accompagne également d'une diminution de l'appréciation découlant de la consommation ou de la fréquentation de ces sources de stimuli. Il s'agit d'un effet particulièrement pervers des dépendances.

On a toutefois tendance à minimiser l'impact résiduel d'une seule décision. Prelec et Bodner ont montré comment la perception par nous-mêmes de nos propres décisions peut

produire des conséquences positives ou négatives très marquées. Ils proposent un modèle d'*autodiagnostic* (*self-signaling*) de la motivation pour expliquer pourquoi les agents hésitent à prendre des décisions sans conséquence négatives, mais qui peuvent s'avérer être les premiers pas vers une mauvaise habitude. Pour Prelec et Bodner,

> « *the model rests on a distinction between two types of reward (utility):* causal *reward that flows directly from the consequences of choice (whether these consequences are immediate or delayed), and* diagnostic *reward, which is the pleasure or pain derived from learning something positive or negative about one's internal state, disposition, ability, or future prospects.* » (2003 : 279)

Dans certaines situations, l'impact hédonique ne proviendrait pas directement des décisions ou des choix effectifs, mais indirectement d'un autodiagnostic. Par exemple, prendre un verre pendant la matinée est un signe d'alcoolisme, choisir de faire un exercice physique difficile est un signe de santé, consommer de la pornographie est un signe de problème sexuel dans son couple, acheter un item dispendieux à crédit est le signe d'une irresponsabilité financière, s'abstenir de voter est le signe d'une irresponsabilité citoyenne, faire un don à un organisme de charité est le signe qu'on est une personne altruiste, etc.

L'autodiagnostic peut être en amont motivé par des considérations que les auteurs classent en trois catégories : intrinsèque, instrumentale et magique (2003 : 281). Les raisons d'un autodiagnostic sont *intrinsèques* quand l'agent se préoccupe davantage des traits ou dispositions pour eux-mêmes que des décisions ou comportements qu'ils engendrent. Je peux, par exemple, refuser d'aller à la chasse non pas parce que je suis contre la pratique de tuer des animaux pour le plaisir, mais simplement parce que je ne veux pas être le genre de personne qui s'adonne régulièrement à cette activité. Les raisons d'un autodiagnostic peuvent être *instrumentales* quand l'agent ne se préoccupe que des conséquences qu'une disposition particulière engendre. Par exemple, le fait d'être une personne qui fume ne me pose pas de problème comme tel, mais si je deviens fumeur j'augmente substantiellement les risques de maladie. Finalement, les raisons d'un autodiagnostic sont *magiques* quand l'agent se préoccupe des dispositions sous-jacentes parce qu'elles seront corrélées aux dispositions d'un bassin de population, et prédiront les comportements de leurs membres. Un exemple de ce genre de raisonnement est la décision d'un agent d'aller voter pour tel candidat parce qu'il estime que cela est le signe que plusieurs autres iront voter pour ce candidat.

Il ne fait pas de doute que l'autodiagnostic est un facteur de contrôle de soi. D'ailleurs, c'est ce qui explique les rechutes des personnes au comportement addictif. Des décisions ou choix ayant à première vue un impact hédonique résiduel négligeable ont un impact beaucoup plus grand s'ils s'accompagnent d'autodiagnostics. Toutefois, cela n'arrive pas dans tous les cas de faiblesse de volonté qui impliquent la réalisation de décisions isolées à impact négligeable. Le modèle de Prelec et Bodner met en relief des conséquences des choix qui avaient jusque-là passé sous le radar. Si une théorie de l'allocation intertemporelle de la faiblesse de volonté est correcte, alors elle ne devrait pas seulement porter sur des décisions isolées, mais également sur des séries de même type.

9.2.3 Akrasie et adaptation hédonique : des schémas d'allocation semblables pour des profils expérientiels semblables?

Les agents prennent beaucoup de décisions qui présentent un profil hédonique décroissant, mais qui ne relèvent pas, à proprement parler, de cas de faiblesse de volonté. C'est le cas de nombreuses décisions de consommation qui impliquent une forme ou une autre d'*adaptation hédonique*[106]. L'adaptation hédonique consiste en une décroissance marquée du profil expérientiel que présente la consommation d'un bien particulier (ex. : une voiture neuve), d'un type de bien (ex. : du chocolat) ou de la fréquentation d'une personne particulière. Obtenir une cuisine neuve, une nouvelle chaîne stéréo, un nouvel ordinateur portable, une nouvelle coupe de cheveux, une nouvelle maison, un nouveau mobilier, produit au début une quantité de plaisir importante qui provient de l'usage, de la contemplation et du rappel imaginaire de ces biens. Mais le plaisir qu'on en retire décroît au fur et à mesure que la consommation s'étend dans le temps. Le nouveau mobilier, pourtant acquis à grands frais, n'est même plus remarqué six mois après. Pour une certain catégorie de biens (et de personnes), leur surconsommation (ou surfréquentation), produit non seulement moins de plaisir marginal, mais renverse à un certain point la qualité de l'expérience hédonique. Le meilleur gâteau au caramel devient rapidement écœurant, le meilleur ami devient ennuyeux lorsqu'on le côtoie à longueur de journée. Pour ces cas, il

[106] L'adaptation hédonique est le pendant psychologique du phénomène économique de l'utilité marginale décroissante. La notion d'utilité marginale décroissante fut initialement introduite pour expliquer pourquoi des biens pourtant essentiels comme l'eau potable valent moins cher sur le marché que des biens accessoires comme les biens de luxe.

existe un point de satiété au-delà duquel la consommation et la fréquentation présentent un rendement marginal négatif.

Cela dit, l'adaptation hédonique n'est pas, comme telle, fonction du nombre de points de consommation ou de fréquentation. Une adaptation hédonique ne se produit que lorsque les points de consommation ou de fréquentation d'un bien, d'un type de bien ou d'une personne sont temporellement rapprochés. Je suis content de voir mon meilleur ami à condition que nos rencontres soient suffisamment espacées. C'est ce qui justifierait d'ailleurs rationnellement l'injonction aristotélicienne de rechercher la modération en toute chose[107].

Bien qu'ils présentent des profiles expérientiels (décroisant) semblables, on peut distinguer les décisions akratiques des décisions de consommation ou de fréquentation qui impliquent une adaptation hédonique. L'adaptation hédonique est le processus psychologique qui sous-tend le rendement hédonique décroissant de certaines de nos décisions (de consommation ou fréquentation) sous certaines conditions (de rapprochement temporel de point de consommation). Il ne s'agit pas, à proprement parler, de décision.

Beaucoup de décisions akratiques impliquent l'activation d'un mécanisme d'adaptation hédonique. C'est souvent le cas pour les comportements de surconsommation. Mais cela n'est pas toujours le cas. Je peux, par exemple, succomber à la tentation d'acheter un mobilier de luxe largement au-dessus de mes moyens. Or, comme ce mobilier est destiné à meubler mon chalet, et que je ne le visite pas souvent, le plaisir découlant de l'usage et de la contemplation du mobilier sera renouvelé à chaque fois. Pourtant, en décidant d'acheter ce mobilier à grand prix (décision akratique), j'opte pour une distribution de conséquences qui concentre des expériences hédoniques positives (ex. : plaisir d'acquérir un bien de luxe, sentiment de distinction ou de réussite sociale, mettre fin à la pression interne exercée pas les mécanismes de contrôle de soi, etc.) dans l'immédiat, et négatives dans un futur plus éloigné. Or, ces dernières conséquences *ne découlent pas directement de ma consommation du bien dans le futur* (ex. : usage, contemplation, etc.), mais simplement des coûts

[107] Il s'agit d'une injonction qu'on doit cependant raffiner, étant donné que les biens comportent des aspects positifs, dont la consommation continue à profiter à l'agent, e*t* des aspects dont la consommation accentue le caractère négatif. La consommation de bière, par exemple, peut élargir notre champ de comparaison et raffiner notre goût, mais également nous faire prendre des kilos en trop. L'évaluation synthétique des divers aspects positifs et négatifs de la consommation à répétition d'un certain bien fut théorisée de manière succincte et élégante par Coombs et Avrunin (1977 : 261-266).

d'opportunités que ma décision initiale a engendrés (ex. : manque d'argent, privation excessive, etc.).

Une autre manière de distinguer les décisions akratiques des décisions produisant une adaptation hédonique est le fait qu'il existe un autre type d'adaptation hédonique qui ne présente pas un profil décroissant. On ne s'adapte pas seulement aux bonnes choses, mais (heureusement!) également aux mauvaises choses. S'acquitter d'une besogne devient plus doux au fur et à mesure que nous avançons. Aussi, on obtient un profil expérientiel croissant lorsqu'on décide de s'acquitter d'une besogne d'un seul trait[108]. Les décisions d'allocation intertemporelle qui impliquent une adaptation hédonique ne présentent donc pas toujours le même profil expérientiel que la faiblesse de volonté. La décision répétée de subir des interventions douloureuses rapprochées dans le temps a pour effet d'aplanir le profil hédonique dans ces interventions.

Il est toutefois indéniable que les décisions akratiques impliquent souvent la réalisation d'une adaptation hédonique. L'absence de modération dans la consommation de nourriture et de biens de luxes, laquelle présentent un potentiel très marqué d'adaptation hédonique, relève souvent de cas de faiblesse de volonté. Mais l'adaptation hédonique ne fait, dans ces cas, qu'accentuer le caractère décroissant du profil hédonique de la décision akratique.

9.2.3.1 Schémas d'allocation semblables ne relevant pas de l'akrasie et nécessité d'une théorie de la motivation

Il nous arrive constamment de prendre de mauvaises décisions de consommation. La plupart du temps, ces décisions présentent un profil expérientiel décroissant – sans quoi il ne s'agirait pas de mauvaises décisions. Je peux m'acheter une automobile usagée que j'obtiens pour un bon prix, mais qui s'avérera rapidement être une source de problèmes importants. Je peux m'acquitter trop rapidement d'une dette qui me laissera dans une situation financière précaire. Je peux décider de commander un plat au menu d'un restaurant coûteux, mais qui s'avérera peu satisfaisant. Dans les cas où je prends une décision et m'en mords ensuite les doigts, seulement certains d'entre eux relèvent d'épisodes de faiblesse de volonté.

[108] On peut utiliser à son profit le mécanisme d'adaptation hédonique en s'abstenant de prendre des pauses lorsqu'on accomplit des activités déplaisantes et en s'en imposant lorsqu'on accomplit des activités plaisantes (Ariely, 2010b).

C'est pour cette raison que la description de la structure temporelle des choix d'allocation ne permet pas d'identifier l'ensemble des conditions nécessaires et suffisantes des décisions akratiques. L'examen des allocations nous permet d'identifier les conditions nécessaires des cas de faiblesse de volonté. On doit également examiner la nature des motivations qui poussent les agents à faire des choix d'allocations au profil décroissant pour être en mesure de formuler un critère opérationnel.

9.3 Conclusion

Les propriétés temporelles des distributions des conséquences sont des caractéristiques centrales de la faiblesse de volonté, autant dans sa forme synchronique que diachronique. L'ignorance de ce fait conduit à surestimer les différences entre les cas de stricte akrasie et les cas d'akrasie tout court. Or, les cas d'irrésolution, d'addiction, de négligence, de non-observance, de procrastination, d'impatience, de surconsommation compulsive, de non-assiduité, de gloutonnerie, de compulsivité ou d'apathie présentent tous un profil décroissant, alors qu'une option présentant un meilleur profil est à la portée de l'agent.

Le choix d'une allocation décroissante (ou insuffisamment croissante) ne constitue pas cependant une condition suffisante pour diagnostiquer un épisode de faiblesse de volonté. Rares sont les contextes de choix non triviaux où il est, pour l'agent, facile d'identifier le choix optimal[109]. L'agent utilise souvent des heuristiques pour découvrir des choix satisfaisants alors que des choix optimaux sont disponibles. Il ne s'ensuit pas qu'il vît alors un épisode de faiblesse de volonté. Comme je vais le montrer dans le prochain chapitre, tout dépend des raisons qui le motivent à opter pour une allocation désavantageuse.

[109] Voir chapitre 7.

10

La théorie du choix intertemporelle (2e partie) : les antécédents motivationnels des décisions akratiques

La sagesse, c'est la prévoyance.

–Térence

10.1 Une théorie de la motivation pour les choix intertemporels

Une analyse conceptuelle satisfaisante de la faiblesse de volonté devait, selon moi, comporter essentiellement deux chapitres importants traitant des deux concepts fondamentaux nécessaires pour comprendre le phénomène. Un chapitre portant sur les *choix* et un autre sur la *motivation* à faire ces choix.

Historiquement, les philosophes ne se sont intéressés qu'aux motivations qui poussent les agents à prendre des décisions akratiques et très peu aux décisions elles-mêmes – d'ailleurs souvent confondues avec les actions ou le comportement. Il en résulte une compréhension appauvrie du phénomène, même concernant notre compréhension de la motivation.

Comme je l'ai montré au précédent chapitre, identifier les propriétés pertinentes des choix akratiques ne mène cependant pas à la formulation d'un critère pour la faiblesse de volonté. Il s'agit de conditions nécessaires, mais non suffisantes. Une analyse compréhensive et détaillée des motivations à faire de tels choix devrait fournir la pièce manquante pour établir un critère.

Pour ce faire, on ne devrait pas mettre l'emphase uniquement sur le jugement de l'agent et ses paramètres cognitifs – comme ce fut le cas pour une bonne partie de la tradition philosophique –, mais identifier les paramètres désidératifs (désirs, préférences, volitions), émotifs (plaisir et souffrance d'anticipation) et normatifs (heuristiques et règles décisionnelles) des procédures expliquant les choix akratiques. Paradoxalement, c'est parce qu'on a porté initialement attention aux propriétés des choix qu'on en arrive à une conception plus riche de la motivation. Si l'on ne se contente que d'examiner les

motivations, on en dressera un portrait incomplet. D'ailleurs, ce qu'on peut reprocher aux analyses qu'on trouve dans la littérature philosophique est qu'elles mettent exclusivement l'accent sur le fait qu'une décision akratique est *inconsistante* soit avec un jugement (toutes-choses-considérées, inconditionnel relatif, etc.) qu'on a *antérieurement formé*, soit avec une décision (intention, plan, résolutions, etc.) qu'on a *antérieurement prise*. Ce faisant, on néglige le fait que les agents prennent dans une très large proportion des décisions pour améliorer leur sort *dans l'avenir*, et que, par conséquent, la clef de la faiblesse de volonté, ainsi que son caractère irrationnel, réside en bonne partie dans les *motivations prospectives* de l'agent.

Du reste, à défaut d'avoir une conception claire de la nature des choix que les agents akratiques font et de ceux qui leur sont disponibles, il devient difficile de voir en quoi la faiblesse de volonté est irrationnelle. Si l'on n'a pas pris le temps d'identifier les propriétés de ces choix, alors on ne peut pas apprécier clairement leurs écueils et la mesure de leur échec. Et si tel est le cas, alors on ne voit que de manière confuse l'irrationalité de ces choix.

10.1.1 Motivation à choisir une allocation intertemporelle : les attitudes prospectives

Les choix d'allocation intertemporelle typique de la faiblesse de volonté sont-ils motivés en amont par des considérations distinctes des autres choix d'allocation au profil décroissant? Sur la base des récents résultats de la psychologie, de la neurologie et de l'économie expérimentale, on peut estimer que c'est le cas. Toute la littérature qui porte d'une manière ou d'une autre sur les *attitudes prospectives* des agents, c'est-à-dire la manière dont ils envisagent le futur, fournit des clefs qui nous permettent de solutionner certains problèmes caractéristiques de l'akrasie (ex. : l'instabilité dynamique des préférences, le regret subséquent, la motivation à prendre des résolutions, la procrastination, l'impatience, etc.) et de fournir un portrait plus complet du phénomène.

Le fait que certains organismes puissent anticiper l'avenir leur permet plus aisément de moduler leurs allocations et de les réviser s'il y a lieu. C'est notamment le cas des humains. Pour les autres, les mécanismes d'allocation sont plutôt rigides. Ils sont déterminés par des algorithmes plus ou moins simples et ne permettent pas toujours une révision des allocations inadéquates. Mais la capacité que les organismes, comme nous, ont de

« produire du futur », pour reprendre une expression de Dennett (1996), au moyen de dispositifs cognitifs et émotionnels, comme l'imagination ainsi que des artefacts culturels (ex. : support d'information au sujet des dangers liés à l'usage du tabac), permet de faire des choix allocatifs très sophistiqués et très étalés dans le temps – du moins pour ce qui est des humains.

Les attitudes prospectives qui motivent les choix d'allocation ont les mêmes paramètres que les motivations pratiques en général. Choisir une allocation dépend évidemment de considérations désidératives : les goûts et préférences pour certaines conséquences ou moyens pour les produire, la volonté de réaliser un résultat, etc. Choisir une allocation dépend en partie de considérations cognitives : la probabilité estimée de la réalisation de certaines conséquences, la visualisation de l'ordre de réalisation d'étapes pour obtenir un résultat visé, l'estimation de la qualité, la durée et l'intensité de l'expérience hédonique découlant de l'obtention de ce résultat (*profil hédonique anticipé*), etc. Le choix dépend également de considérations émotives : les humeurs, les « feelings », les sentiments et les appréciations subjectives actuels découlant de l'anticipation d'événements futurs. Et, finalement, le choix dépend de considérations normatives : heuristiques décisionnelles, règles d'évaluation des allocations, méthode d'optimisation, etc.

Il n'est pas exclu que les attitudes prospectives soient motivées en amont par d'autres attitudes qui ne sont pas à proprement parler des attitudes prospectives. Les considérations cognitives portant sur la réalisation d'un événement futur reposent la plupart du temps sur les données de la mémoire d'observation de relations causales passées, sur des liens d'inférence atemporels (obtenus par déduction, induction, inférence à la meilleure explication, etc.), sur des dispositions et des théories innées portant sur la structure physique du monde (Baillargeon, 1987) ainsi que sur les états mentaux des autres agents. Que les événements du passé déterminent souvent les choix pour l'avenir est un fait trivial qui ne contredit pas l'idée que les attitudes motivant l'agent *en première instance* sont essentiellement de nature prospective. Lorsqu'on demande à un agent d'expliquer pourquoi il a pris telle ou telle décision d'allocation, c'est habituellement par la mention du contenu de l'un de ces paramètres qu'il nous répond. Cela n'implique pas cependant que ces contenus soient tous conscients. Les agents ont souvent beaucoup de mal à expliciter les règles décisionnelles qu'ils utilisent, et des résultats en neuropsychologie montrent

clairement qu'ils ne sont pas toujours conscients des désirs qui les motivent et de leur appréciation hédonique des conséquences (Nisbett & Wilson, 1977 ; Berridge & Robinson, 2003). Mais, acceptant l'hypothèse que les agents prennent des décisions la plupart du temps en fonction de l'avenir, et c'est clairement le cas, je crois que, pour expliquer les décisions akratiques, on doit mettre l'accent sur les attitudes prospectives[110].

Les attitudes prospectives peuvent être plus ou moins sophistiquées. Elles deviennent sophistiquées lorsqu'elles comportent des considérations métacognitives, méta-désidératives ou méta-affectives. L'examen de ces méta-attitudes peut révéler beaucoup de choses sur la faiblesse de volonté, autant du côté de ses facteurs de réalisation que du côté de ses facteurs de résistance. Par exemple, un agent peut faire un choix d'allocation typique de la faiblesse de volonté parce qu'il veut rester spontané ou ne pas devenir le genre de personne qui se donne des contraintes, ou décider de ne pas choisir ce type d'allocation parce qu'il a peur d'avoir des regrets subséquents ou parce qu'il ne veut pas exacerber un désir malsain qu'il ne veut pas avoir, etc. La psychologie, les neurosciences et l'économie comportementale n'ont que peu ou pas du tout investigué du côté de ces méta-attitudes. On doit se rabattre sur les descriptions fines des littéraires et romanciers pour avoir un aperçu de leur nature et du rôle qu'elles peuvent jouer dans l'explication de la faiblesse de volonté et des mécanismes de contrôle de soi. Aussi, je les négligerai dans une large mesure dans l'exposé des motivations à faire des choix akratiques.

10.1.2 Attitudes prospectives : escompte du futur et choix d'allocation au profil décroissant

Les chercheurs œuvrant dans le domaine de la motivation ont tendance à synthétiser l'ensemble des considérations paramétriques dans un seul et unique vecteur motivationnel. Pour les attitudes prospectives, les théoriciens ont introduit la notion de l'*escompte du futur* pour jouer ce rôle[111]. L'escompte du futur représente synthétiquement l'ensemble des

[110] Il pourrait arriver cependant que des épisodes de faiblesse de volonté soient motivés en première instance par des attitudes rétrospectives. Je peux décider de succomber à la tentation de fumer une cigarette, alors que j'avais pris la résolution d'arrêter, pour me punir d'une faute morale passée. Je ne sais pas si ce n'est toutefois qu'une possibilité théorique sans réel fondement empirique.

[111] C'est depuis la publication des travaux fondateurs d'Irving Ficher (1930, 1970) et de Paul Samuelson (1937) que les économistes et les psychologues synthétisent les attitudes prospectives dans un seul et unique taux d'escompte, bien que les premiers travaux sur les attitudes prospectives en économie remontent aux publications de Rae, Senior et Jevon. Ces

considérations orientées vers le futur pertinentes pour choisir une allocation plutôt qu'une autre. En d'autres mots, l'escompte du futur est le résumé synthétique de l'ensemble des raisons qui font que nous nous préoccupons davantage des conséquences réalisables *à certains moments du temps plutôt qu'à d'autres*[112].

D'ailleurs, si la position temporelle des conséquences distribuées est un facteur de choix, pour un agent qui escompte le futur, leur rapprochement graduel du moment présent modifiera par le fait même leur valeur. C'est pour cette raison qu'une description de l'escompte du futur doit mentionner un *taux* d'escompte qui caractérise pour un agent l'ampleur de sa dévaluation des conséquences futures. Théoriquement, le taux d'escompte pour un agent dans le domaine financier correspond *grosso modo* au taux d'intérêt minimal auquel il serait prêt à effectuer des placements. Si je ne suis disposé à me départir maintenant de 100 $ que pour obtenir au moins 130 $ dans un an, alors j'escompte le futur à un taux annuel de 30 %. Par contre, les taux d'escompte ne sont pas toujours aussi faciles à calculer que dans le domaine financier. D'ailleurs, les agents n'escomptent pas le futur de la même manière pour tous les domaines d'activités. On peut escompter le futur à un taux beaucoup plus élevé pour les conséquences découlant, par exemple, de l'ingurgitation de matières grasses que pour les conséquences découlant de l'achat de matériel informatique ludique. Les agents au caractère akratique n'éprouvent pas (dans la plupart des cas) des problèmes de faiblesse de volonté dans tous les domaines d'activité.

Les taux d'escompte variés synthétisent le poids des motivations hétérogènes des agents pour l'obtention d'allocations aux profils hédoniques croissants, décroissants ou uniformes. Un taux d'escompte négatif correspond à une dévaluation du futur, un taux positif, à une surévaluation du futur, et un taux neutre, à une égale considération pour les conséquences réalisables à n'importe quel moment du temps.

---

derniers, insatisfaits du modèle d'Adam Smith, ont tenté d'expliquer la richesse des nations par le désir d'accumulation des citoyens et leur promptitude à différer leurs dépenses.

[112] Bien que s'inscrivant au départ dans un modèle *fondé-sur-les-valeurs* (Théorie de l'Utilité Espérée, Théorie des Perspectives), la notion d'escompte du futur s'avère également utile dans un modèle *fondée-sur-les-raisons*. On peut expliquer pourquoi un agent escompte le futur de telle ou telle façon en décrivant un calcul d'utilité qu'il aurait au préalable fait, ou en décrivant un raisonnement pratique qui inclut des considérations désidératives, cognitives, émotionnelles ou normatives, pertinentes parce qu'elles font la différence dans le choix d'une allocation plutôt qu'une autre.

La manière dont les agents akratiques escomptent le futur fournit une ressource conceptuelle claire et simple pour tenter d'expliquer pourquoi ils optent pour des allocations présentant un profil hédonique décroissant – qu'il y ait présence ou non d'une adaptation hédonique. Dans les cas de faiblesse de volonté, les agents escomptent le futur d'une manière qui trahit leur plus grande préoccupation pour les conséquences plus rapprochées d'eux dans le temps. Lorsqu'un agent choisit d'acheter, par exemple, un climatiseur peu coûteux à l'achat, mais ayant un rendement énergétique médiocre, alors qu'il sait qu'il y en a de légèrement plus coûteux, mais beaucoup plus performants, il exhibe un souci plus grand pour les conséquences à court terme (le coût immédiat vs les coûts futurs). Ici, comme ailleurs, l'escompte du futur est *révélé* par les choix effectifs d'allocation intertemporelle. Dans les cas de faiblesse de la volonté, les agents exhibent de la même manière une préoccupation plus grande pour les conséquences hédoniques temporellement rapprochées. Mais ce n'est que le début de l'enquête. Une analyse fine des motivations qui sous-tendent ces choix doit nous permettre de distinguer les cas d'akrasie des autres décisions. Aussi, cette analyse devrait nous fournir la base nécessaire pour juger l'hypothèse de l'irrationalité présumée de la faiblesse de volonté.

10.1.2.1 L'escompte hyperbolique du futur : un trait caractéristique motivationnel de l'akrasie

Les économistes ont cru pendant longtemps que les agents devaient escompter le futur selon un taux négatif, mais constant en l'absence de modification significative des goûts de l'agent, des opportunités disponibles pour lui et de sa capacité de jouissance. Depuis la publication des premiers travaux sur les choix d'allocation chez les pigeons de Rachlin et Green au début des années 1970, les théoriciens en psychologie ont commencé à mettre en doute le caractère réaliste de la théorie standard qui veut que les agents escomptent le futur en fonction d'un taux négatif constant. Les données expérimentales bien documentées tendent à montrer que certains mammifères escomptent en fait le futur en fonction d'un taux décroissant et non d'un taux négatif constant (Strotz : 1955 ; Chung & Herrnstein, 1967 ; Ainslie, 1974). Un taux constant implique que les préférences intertemporelles sont temporellement consistantes (*time-consistent*), ce qui signifie *grosso modo* que les dernières préférences « confirment » les préférences antérieures (Frederick, Loewenstein, O'Donoghue : 2003). En effet, si une alternative future *x* est préférée à une autre alternative

future *y*, et que l'agent les escompte toutes les deux en fonction d'un taux constant *c*, *y* ne sera jamais préférée à *x* malgré le fait que les valeurs de chacune de ces alternatives croissent au taux *c* au fur et à mesure qu'elles se rapprochent temporellement.

Les premières données empiriques qui tendent à montrer que certains animaux n'escomptent pas le futur de cette manière remontent à l'expérience de Rachlin et Green (1972). Les deux chercheurs ont montré que les pigeons n'escomptent pas du tout les récompenses futures en fonction d'un taux constant. Dans l'expérience, chaque pigeon était confronté à deux leviers pourvus d'un indicateur lumineux respectivement rouge et vert. Les indicateurs s'allumaient en même temps et signalaient qu'une certaine quantité de nourriture devenait disponible en un simple coup de bec. Lorsque l'indicateur rouge s'allumait, le pigeon n'avait qu'à piquer le levier pour obtenir immédiatement une quantité de nourriture pendant 2 secondes. Lorsque l'indicateur vert s'allumait, une plus grande quantité de nourriture devenait disponible pendant 4 secondes, mais nécessitait un temps d'attente de 4 secondes. Après avoir essayé successivement les deux alternatives, les pigeons marquaient une nette préférence pour la récompense immédiate en dépit du fait qu'elle est deux fois plus petite.

Les pigeons manquent à cet égard de contrôle de soi. Rachlin et Green ont ensuite modifié les conditions de l'expérience. Les pigeons impatients devaient maintenant se contenter d'appuyer sur le levier qui offre une grande récompense parce que seul l'indicateur vert s'allumait à répétition pendant une période plus ou moins longue. Les pigeons « apprenaient » de cette manière à devenir plus patients.

Mais la variante la plus significative de l'expérience, et celle qui fit couler le plus d'encre, est celle qui consistait à varier les délais respectifs des récompenses. Les chercheurs se sont rendu compte qu'en ajoutant un délai d'attente toujours plus grand avant que les indicateurs s'allument, on pouvait observer un renversement de préférence chez les pigeons. Plus le délai était long, moins les pigeons avaient tendance à choisir la plus petite récompense. Si l'on ajoutait un délai de 10 secondes pour l'obtention des récompenses – donc 10 secondes pour obtenir la petite récompense et 14 pour obtenir la plus grande – les pigeons appuyaient systématiquement sur le levier vert. Ils exhibaient alors plus de contrôle de soi et se montraient donc moins impulsifs. À partir d'une certaine distance temporelle, ce n'est plus

le délai entre deux récompenses qui est le facteur prédominant de choix, mais la différence dans la quantité relative de celles-ci.

Thaler a observé le même schéma chez des sujets humains auxquels on offrait des récompenses en argent et dont on mesurait l'impact motivationnel des délais d'obtention (Thaler, 1981). On demandait à des personnes – des étudiants pour la plupart – d'imaginer qu'ils ont gagné une somme d'argent à la loterie. On leur demandait ensuite de spécifier des montants d'argent supplémentaires à encaisser et pour lesquels ils seraient disposés à attendre pour recevoir le paiement du lot principal. Les personnes devaient spécifier les montants pour des délais d'un mois, d'un an et de dix ans. Les résultats furent comparables à ceux obtenus avec les pigeons. Les personnes demandaient, *proportionnellement au délai*, plus d'argent pour renoncer à des gains rapprochés d'eux dans le temps que pour des gains plus distants.

L'expérience de Thaler fut reprise et raffinée par plusieurs chercheurs. Les résultats obtenus pointent dans une direction : le fait que les agents escomptent le futur en fonction d'un taux décroissant. Toute une littérature s'est développée au sujet du phénomène que les théoriciens appellent *l'escompte hyperbolique*, en raison de la forme hyperbolique qu'on obtient en décrivant le poids hédonique ou motivationnel d'options concurrentes pour chaque instant d'anticipation[113] (figure 5).

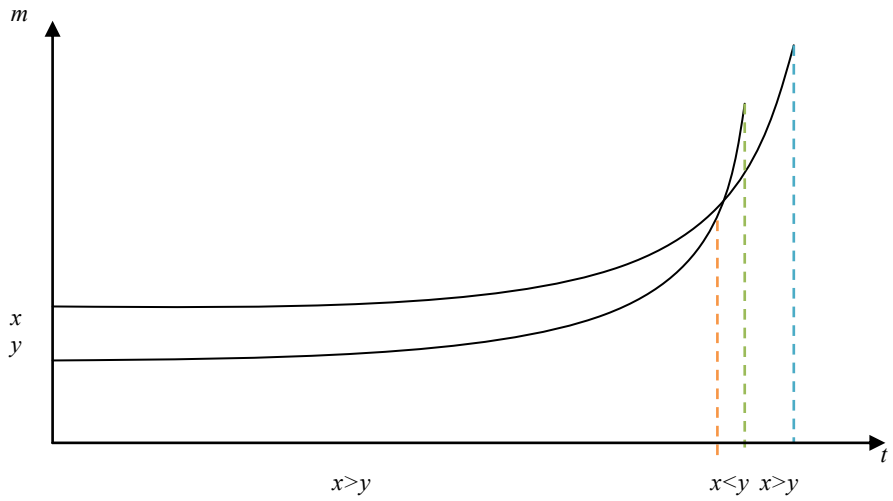

Figure 5

[113] En ce sens, l'escompte hyperbolique est à contraster avec l'escompte exponentiel qui correspond à une courbe décrite par un taux constant.

Avec l'escompte hyperbolique, on voit que des options disponibles ou des conséquences subies à des moments distincts acquièrent un poids motivationnel plus rapidement lorsqu'elles se rapprochent du moment présent. Nous avons tous l'intuition que le poids motivationnel des options ou des conséquences temporellement distantes s'accentue au fur et à mesure que les délais qui nous en séparent rétrécissent. Les émotions que nous ressentons à l'idée de subir des conséquences positives ou négatives qui se rapprochent deviennent de plus en plus vives. Nous sommes sans doute excités à l'idée que nous allons acquérir la maison de nos rêves dans 6 mois ou revoir un être cher après plusieurs années d'absence, mais pas autant que la veille de ces événements. De même, nous pouvons éprouver du déplaisir à l'idée de subir un triple pontage coronarien ou d'annoncer notre intention de rompre une liaison amoureuse à la principale personne intéressée, mais pas autant que la veille de ces événements. Mais la principale découverte que les chercheurs ont faite en décrivant l'escompte hyperbolique est que les conséquences qui sont temporellement plus près de nous gagnent de l'importance beaucoup plus rapidement que les conséquences plus éloignées, même si toutes ces conséquences « avancent vers nous à la même vitesse ». Comme les psychologues George Ainsli et Nick Haslam l'ont judicieusement fait remarquer, le phénomène est analogue à l'illusion d'optique qui survient lorsqu'on marche en direction d'édifices de hauteurs différentes (Ainsli & Haslam, 2003). À mesure que nous nous rapprochons d'eux, l'édifice le plus près semble « grandir » plus rapidement que celui qui est plus éloigné. Aussi, lorsque nous arrivons à un certain point, les deux semblent avoir la même hauteur, et, passé ce point, le plus près semble maintenant plus élevé que le second. Une fois les édifices derrière nous, l'illusion se dissipe (figure 5).

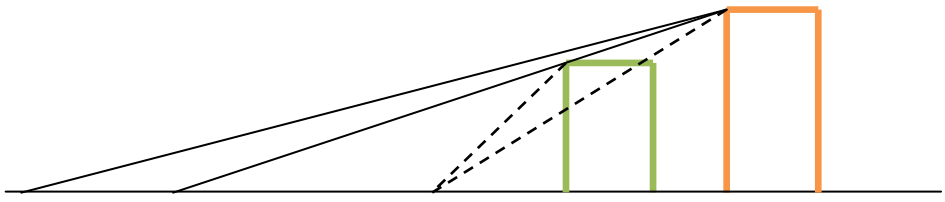

Figure 6

Bien que nous soyons en mesure d'estimer la hauteur relative de chacun des édifices lorsque nous sommes à bonne distance, un effet de perspective peut en venir à distordre momentanément notre estimation. Aussi, nous optons pour des allocations intertemporelles avantageuses lorsque les conséquences sont presque entièrement distribuées dans une période lointaine. Par exemple, il est plus facile de prendre la décision de subir une intervention dentaire lorsque cette dernière a lieu dans un mois, que lorsqu'elle a lieu le lendemain. Si je compare l'option de m'y soumettre avec l'option de m'en abstenir, j'aurai une vue d'ensemble plus conforme à mon intérêt bien pesé si les conséquences qu'elles engendrent ont lieu dans un avenir lointain. Aussi, je serai alors disposé à prendre un rendez-vous chez le dentiste. Mais comme le rapprochement temporel des conséquences anticipées en distordra mon évaluation, je deviendrai davantage disposé à annuler mon rendez-vous. Si je le fais, j'éprouverai typiquement des regrets une fois la date de mon intervention passée, car j'aurai alors recouvré une vision claire de mon intérêt bien pesé. L'analogie avec les illusions d'optique est éclairante, mais on ne doit pas la pousser trop loin. On n'est plus leurré par une illusion quand on en connaît, ne serait-ce que sommairement, les mécanismes. Par contre, le fait de savoir que nous escomptons hyperboliquement le futur ne nous met pas à l'abri des erreurs souvent graves que nous sommes disposés à faire en choisissant certaines allocations. L'escompte hyperbolique nous pousse de manière plus ou moins irrésistible à vivre des épisodes de faiblesse de volonté. Les illusions sont sources d'étonnement parce qu'elles ne concordent pas avec les informations auxiliaires qui nous orientent dans d'autres directions. Mais la plupart d'entre nous connaissent mal la manière dont nous escomptons le futur. Nous pouvons sans peine identifier la plupart des attitudes prospectives qui nous motivent à faire tel ou tel choix, mais nous n'avons pas une vue d'ensemble de la dynamique de ces attitudes. Aussi, nous sommes souvent disposés à nous réabonner à la salle d'entraînement, à refaire une diète, à nous promettre de remettre cette session tous nos travaux dans les délais, à ramasser les feuilles mortes cet automne avant les premières chutes de neige, etc., et ce, bien que nous nous soyons fait les mêmes promesses et ayons pris les mêmes résolutions dans le passé, et que nous ne les ayons pas respectées. C'est comme si nous n'apprenions pas de nos erreurs passées parce que nous en ignorions la cause globale, et que nous ne cherchions qu'à identifier des facteurs périphériques lorsqu'on nous interroge sur le sujet.

L'escompte hyperbolique permet d'expliquer pourquoi les agents akratiques optent pour des allocations qui présentent des profils hédoniques inadéquats ou peu avantageux. C'est en quelque sorte un effet de perspective temporelle qui motive les agents à procrastiner et à adopter des comportements impatients. Mais les recherches auxquelles a donné lieu la découverte de l'escompte hyperbolique ont permis d'en identifier plusieurs paramètres importants.

Un des paramètres importants est *la qualité des conséquences anticipées*. Il semblerait que les agents escomptent de manière plus accentuée les conséquences positives que les conséquences négatives (*Sign Effect*). Les personnes sont disposées à payer moins maintenant pour retarder le paiement futur d'un bien qu'ils viennent d'obtenir qu'à payer maintenant pour obtenir un bien qui leur est promis dans l'avenir (Thaler : 1981). Cela explique pourquoi les activités qui présentent à la fois des attributs négatifs et positifs très marqués sont plus souvent l'objet d'épisodes de faiblesse de volonté. Pour un saut en parachute, par exemple, on escomptera les risques de mort d'une manière bien plus marquée que des conséquences comme l'ivresse de la sensation de voler.

Un autre paramètre est celui de *magnitude* (*Magnitude Effect*). Les agents escomptent moins les conséquences importantes que les conséquences plus négligeables. Les personnes qui sont indifférentes à 15 $ aujourd'hui et à 60 $ dans un an seront en moyenne également indifférentes à 25 $ aujourd'hui et à 35 $ dans un an, à 3000 $ aujourd'hui et à 4000 $ dans un an (Thaler, 1981). En fait, plus les conséquences sont importantes pour nous, plus nous serons disposés à nous montrer patients.

Le paramètre de l'asymétrie des *délais de livraison et de l'accélération de livraison* (*Delay –speedup asymmetry*) permet aussi de déterminer le taux d'escompte d'un agent. Il semble que nous soyons plus disposés à payer pour accélérer la réalisation d'une conséquence positive prochaine dont on pensait initialement la réalisation immédiate, qu'à payer pour obtenir immédiatement cette réalisation alors qu'on pensait qu'elle aurait lieu dans le futur (Loewenstein, 1988). Il semble que le coût de renoncement pour un gain soit en partie déterminé par un point de référence temporel, fixé par une attente initiale, et à partir de duquel l'agent escomptera un gain futur.

Cela dit, l'escompte hyperbolique fournit un élément descriptif de la motivation essentielle pour rendre compte des cas d'irrésolution et de non-observance – comme ceux sur lesquels des théoriciens comme Bratman et Holton mettent l'accent. Comme nous avons une vision davantage conforme à notre intérêt bien pesé lorsque les conséquences des choix sont plus éloignées dans le temps, nous sommes par conséquent plus disposés à prendre de bonnes décisions pour l'avenir lorsque nous avons suffisamment de recul. Comme il a été mentionné au chapitre 7, nous prenons des décisions pour l'avenir (intentions, plans et résolutions) parce qu'il s'agit de bons moyens cognitifs pour atteindre des objectifs à plus ou moins long terme. Ce que la conception de Holton passe sous silence est le fait significatif que les conditions sous lesquelles l'agent planifie de faire quelque chose ne sont pas les mêmes que les conditions sous lesquelles il s'apprête à exécuter son plan. La structure des incitations a changé même si la structure de l'allocation intertemporelle demeure la même.

Une théorie de l'allocation intertemporelle comprenant un chapitre sur l'escompte hyperbolique permet d'aller plus loin et elle constitue un complément essentiel à des théories comme celle de Holton. Mais elle permet de rendre compte de cas de faiblesse de volonté qui ne comportent pas de résolutions antérieures, comme les cas de négligence et d'habitudes nocives non diagnostiquées par l'agent lui-même. Les agents qui ont un problème d'akrasie chronique *peuvent être ou non* des agents qui prennent systématiquement des décisions pour le futur qui reflètent leur intérêt bien pesé. Du moment qu'ils escomptent le futur de manière hyperbolique, ils sont sujets à vivre des épisodes de faiblesse de volonté.

Cela dit, les considérations sur les résolutions et les plans nous laissent cependant croire que l'escompte hyperbolique concerne exclusivement les attitudes prospectives que les agents entretiennent à l'égard des conséquences sur le long terme. Aussi, il est naturel de dire qu'escompter le futur de manière hyperbolique motive les agents à satisfaire davantage leurs intérêts à court terme (ou maintenant) que leurs intérêts à long terme (ou plus tard). Par conséquent, il peut être commode de dire que l'agent akratique est tout simplement celui qui choisit de favoriser ses intérêts à court terme alors qu'il ne devrait pas le faire. C'est une caractérisation essentiellement correcte, mais à la condition qu'on offre des précisions sur les catégories du court et du long terme.

On utilise souvent la notion de long terme pour des laps de temps de plusieurs mois, voire plusieurs années. Or, ça ne s'applique pas aux décisions qui consistent à manger son dessert avant son repas principal ou à prendre un verre de trop. Il est plus judicieux de considérer la distinction comme une distinction fonctionnelle et non catégorielle. En tant que distinction fonctionnelle, le court terme (ou le « maintenant ») est défini contextuellement à partir de l'intervalle de temps le plus petit nécessaire pour résoudre un problème pratique dans un domaine spécifique. Tandis que le long terme (ou le «plus tard») est défini à partir de l'intervalle le plus long pour les résoudre. Par exemple, dans le contexte d'un jeu vidéo, le court terme est souvent en deçà d'une seconde et le long terme de quelques minutes tout au plus. Par contre, dans le contexte d'un plan de vie, le court terme s'étend sur plusieurs mois et le long terme sur plusieurs années.

La version fonctionnelle de la distinction est adéquate pour décrire correctement les épisodes de faiblesse de volonté. Elle permet de rendre compte de ce que Ainslie a décrit comme des cycles d'impatience typique des épisodes de faiblesse de volonté (2001 : 64). Un cycle complet coïncide avec un profil hédonique qui débute par un accroissement rapide de l'expérience positive qui, après avoir atteint un pic maximal décroît plus ou moins rapidement et déchoit dans une longue période de déplaisir après avoir franchi un seuil hédoniquement neutre. On peut décrire ce cycle de manière non technique : on est tenté, on succombe et on le regrette ensuite amèrement.

Ainslie classe les compulsions dans les cycles qui s'étendent sur plusieurs mois ou années. Le travail compulsif, les attitudes socialement condamnables comme le mépris et la pédanterie en sont des exemples. Les dépendances ont des cycles s'étendant sur des heures, voire des jours. Elles concernent le jeu, la consommation de matériel pornographique, les substances ou drogues, les attitudes colériques et promptes au reproche. Les envies et les appétits enclenchent des cycles qui s'étalent sur des minutes et des heures. Ils induisent des comportements comme le vagabondage sur Internet et les postes de télévision, la surconsommation de nourriture, la paresse. Les démangeaisons et les tics ont aussi le même profil hédonique escompté, mais elles engendrent des cycles beaucoup plus courts, ayant une durée de quelques secondes. Étant donné la durée de ces cycles, les périodes pendant lesquelles les agents seront disposés à reconnaître qu'ils ont agi de manière irrationnelle et

en éprouver des regrets se dérouleront quelques secondes après le début du cycle, quelques années après dans le cas des cycles très longs.

L'idée d'inclure les démangeaisons, les tics et autres troubles obsessionnels ayant un cycle très court ainsi que les dépendances dans la catégorie des épisodes de faiblesse de volonté ne cadre pas tout à fait avec ce qu'on trouve dans la littérature philosophique. À quelques exceptions près, les philosophes contemporains excluent les comportements irrésistibles des épisodes de faiblesse de volonté parce qu'ils ne sont pas accomplis ou décidés de manière libre. Dans la conception standard, un agent qui vit un épisode de faiblesse de volonté *aurait pu* agir ou décider autrement. Cela peut sembler représenter un problème pour une conception plus générale de la faiblesse de volonté qui reposerait sur l'escompte hyperbolique et des allocations intertemporelles. On pourrait avoir des cas d'allocation qui présentent un profil hédonique identique au profile *A* (figure 2) et dont les attitudes d'anticipation peuvent être décrites comme un cas d'escompte hyperbolique, mais ne pas relever d'un cas de faiblesse de volonté parce que la condition de liberté de choix n'est pas respectée. Les démangeaisons, les tics et les cas de dépendance sévères seraient de cette manière exclus des cas akratiques[114]. Cependant, on peut arguer que tout ceci dépend de ce qu'on veut dire par « *être capable de…* ». Si résister à un tic nerveux est de même ampleur que la tâche de courir le 100 mètres en moins de 9 secondes, alors on ne sera pas plus faible en échouant à le contrôler qu'en échouant à courir le 100 mètres en moins de 9 secondes. Mais si cela signifie que ce n'est pas hors de portée, ou qu'avec un entraînement on y arriverait dans des délais raisonnables, même en tenant compte de ses prédispositions et de ses limitations génétiques, alors on peut classer ces comportements dans les cas de faiblesse de volonté. En fait, rien ne nous permet de conclure *a priori* que de résister à un tic est une tâche hors de notre portée comme courir le 100 mètres en moins de 9 secondes l'est. Les thérapies cognitivo-comportementales conçues pour les sujets atteints de TOC (troubles obsessionnels compulsifs) visent justement à établir les bases d'une capacité de contrôle de soi permettant d'endiguer ou d'ignorer certaines motivations viscérales.

[114] Watson (1977) a critiqué en détail l'idée que ce qui différentiel les épisodes de faiblesse de volonté authentiques des comportements irrésistibles. Il soutient que la faiblesse de volonté consiste toujours en un échec ne nos mécanismes de contrôle de soi et de l'absence de tels mécanismes.

10.1.2.2 Comment expliquer que les agents escomptent hyperboliquement le futur?

Bien que représentant une découverte importante, l'escompte hyperbolique ne constitue pas comme telle une explication des cas de faiblesse de volonté. Il s'agit plutôt d'une description dynamique (obtenue en comparant les attitudes prospectives à l'égard des événements qui se rapprochent dans le temps) et synthétique (sans donner de détails sur les motifs sous-jacents) des raisons qui motivent les agents à choisir certaines allocations globalement moins avantageuses. Une analyse des raisons qui motivent les agents à *dévaluer* le futur comme ils le font lorsqu'ils l'escomptent hyperboliquement devrait révéler les motivations fondamentales à prendre des décisions akratiques. Cela nous amène à poser deux questions : (*a*) pourquoi les agents akratiques dévaluent le futur? ; et (*b*) pourquoi ils le dévaluent de manière hyperbolique?

(*a*) En fait, on peut identifier plusieurs motifs, de nature hétérogène, pour dévaluer le futur et opter pour une allocation au profil décroissant, alors que des allocations moins décroissantes, homogènes ou croissantes sont disponibles. La plupart de ces motifs n'ont intuitivement rien à voir avec la faiblesse de volonté et constituent même à l'occasion de très bonnes raisons pour faire des choix. Examinons quelques-unes de ces raisons :

1/ **Limites mémorielles** : Une pratique courante dans le commerce au détail consiste à offrir des biens à un coût initial un peu plus élevé, mais avec la possibilité pour le client de recevoir un rabais postal relativement substantiel. Il est cependant de la responsabilité du client de remplir le formulaire approprié et de le poster au fabricant. Seulement, l'achat n'est avantageux que dans la mesure où l'on se rappellera en temps et lieu d'effectuer la requête. Étant donné les risques d'oubli, il peut être rationnel de s'abstenir de choisir l'option la plus avantageuse à long terme.

2/ **Coûts d'opportunité** : On peut refuser d'investir des ressources en vue d'une jouissance lointaine parce que cela implique un coût d'opportunité actuel somme tout peu important, mais qui ferme en revanche des portes pour de meilleurs investissements potentiels. Le coût d'opportunité représente un motif circonstanciel, mais réel pour dévaluer momentanément le futur.

3/ **Probabilités et aversion aux risques** : Si l'on se soucie davantage des conséquences temporellement plus rapprochées, cela peut être parce que la réalisation des conséquences

plus éloignées dans le temps comporte trop d'incertitude. Les agents qui ont une aversion au risque même modérée préfèreront dans beaucoup de situations *un tiens* que *deux tu l'auras*. Aussi, il peut être judicieux de remettre des tâches ennuyeuses à plus tard et de dépenser des montants d'argent importants au lieu de les investir si l'on n'est pas sûr qu'on sera encore en vie demain. Le fait de considérer moins l'avenir que le présent n'est évidemment pas toujours irrationnel. Les probabilités d'interférence que représentent les facteurs externes dans la poursuite de nos objectifs militent en faveur d'une dévaluation de l'avenir. Mais c'est le cas aussi de notre aversion subjective au risque, en tant que paramètre désidératif. Une aversion importante à l'égard du risque motivera le choix d'allocation avec un profil décroissant très marqué. Les aversions au risque trop importantes, de même qu'une surestimation des probabilités d'échec, pousseront évidemment les agents à faire des choix irrationnels, mais ils ne seront pas akratiques pour autant.

4/ **Changement de goût et de capacité de jouissance** : Nous pouvons être plus ou moins certains de la réalisation d'une conséquence future, mais très incertains de la manière dont elle nous affectera. Je peux, à l'âge de 65 ans, entreposer un vin millésimé qui sera à point lorsque j'aurai 75 ans, mais comme je ne suis pas convaincu d'être capable d'en jouir pleinement, je le consommerai avant. Il en est de même, en ce qui concerne l'incertitude quant au changement futur, de nos goûts pour nous motiver à consommer plus rapidement certains biens ou pour nous adonner à certaines activités qui ne nous plairont hypothétiquement qu'à l'intérieur d'un horizon temporel limité. Dans ces circonstances, le changement de goût peut nous motiver à diminuer l'ampleur des sacrifices que nous étions pourtant prêts à faire pour nous assurer un meilleur avenir. Par exemple, je peux investir une partie de mes économies dans un régime de retraite en fonction d'un profil de dépense que j'escompte avoir après 65 ans. La quantité de ressources dont je devrai me priver maintenant et sur une base régulière dépend de mon espérance de vie et de la consommation que je juge, aujourd'hui, être satisfaisante dans un avenir lointain. Si je pense que je vais m'adonner, pendant les 15 années qu'il me restera à vivre, à des activités de loisir coûteuses comme jouer au golf et voyager, et non à du bénévolat dans ma collectivité, alors je devrai restreindre davantage ma consommation actuelle. Mais, à défaut de connaître avec précision la nature de mes goûts futurs, il est prudent de réviser à la

baisse certains projets d'investissement. Mais si cela se traduit par une dévaluation du futur, il ne s'agit pas intuitivement de cas de faiblesse de volonté[115].

Les 4 types de considérations que je viens de mentionner expliquent pourquoi les agents dévaluent le futur dans certaines situations sans que cela ne relève de cas de faiblesse volonté. Les limites mémorielles, les probabilités de réalisation d'un événement et les coûts d'opportunité relèvent du paramètre motivationnel cognitif. L'aversion au risque, le changement de goût et la capacité de jouissance relèvent des paramètres désidératifs et affectifs. Ces considérations hétérogènes ne motivent pas toujours l'agent à faire des choix rationnels, mais restent *grosso modo* conformes à certains canons de la rationalité pratique (ex. : toujours tenir compte de ses limites cognitives lorsque vient le temps de faire un choix qui implique des conséquences sur le long terme).

(*b*) D'ailleurs, ces considérations peuvent expliquer pourquoi les agents peuvent dévaluer le futur, mais pas pourquoi ils le dévaluent de manière hyperbolique. Elles n'expliquent pas pourquoi l'approche temporelle de conséquences est fortement corrélée avec une inflation croissante, et non avec une inflation fixe.

Pour expliquer pourquoi les agents dévaluent le futur de manière hyperbolique, beaucoup de théoriciens ont inféré l'existence de préférences temporelles non dérivées (*pure time preference*). Si nous escomptons hyperboliquement le futur, alors nous avons une ***aversion pour les délais de gratification*** et/ou une ***préférence pour les délais de frustration*** au sein de notre espace motivationnel. Aussi, l'élément qui expliquerait l'escompte hyperbolique résiderait dans notre paramètre désidératif, et non dans les paramètres cognitifs, affectifs ou normatifs de notre motivation. Mais si chacun des deux types de préférences temporelles constitue à lui seul une condition nécessaire et suffisante pour un escompte hyperbolique, ceux-ci ne jouent pas le même rôle dans l'explication de comportement akratique. L'aversion pour les délais de gratifications expliquerait mieux les comportements impatients, tandis que les préférences pour les délais de frustration, la procrastination. En

[115] La difficulté d'anticiper des changements substantiels dans nos goûts peut néanmoins favoriser la production de décisions akratiques. Loewentein (2005) a mis en relief le fait que les agents sous-estiment l'impact des motivations viscérales (ex. : excitation, curiosité, enthousiasme, etc.) sur leurs goûts et leurs comportements futurs (*hot/cool empathy gap*). Aussi, ils négligeront de prendre des mesures profilactiques de contrôle de soi, alors que ces motivations les rendront disposés à choisir des allocations aux profils nettement décroissants, voire à accomplir des actes moralement répugnants. Mais la difficulté d'anticiper les changements majeurs dans mes goûts n'est pas ce qui explique pourquoi je choisirais, dans l'avenir, une allocation au profil décroissant.

fait, on doit examiner le profil expérientiel des allocations en concurrence. Si les allocations se font concurrence au niveau des instants où seront réalisées les conséquences positives les plus saillantes, alors ce sera l'aversion aux délais de gratification qui fera la différence (ex. : choisir entre des activités plaisantes). Mais si c'est au niveau des conséquences négatives, ce sera la préférence pour les délais de frustration (ex. : choisir entre des besognes). Il y a évidemment beaucoup de cas mixtes – notamment les cas où l'on se montre impatient en procrastinant (ex. : choisir de faire la fête pour s'éviter de rédiger un travail important).

Poser l'hypothèse de préférences temporelles fournit une explication simple et élégante. À l'égard de la structure dynamique étrange de l'escompte hyperbolique, il est de surcroit assez naturel de postuler des préférences qui ne portent pas sur des conséquences comme telles, mais sur leurs instants de réalisation. D'ailleurs, il ne s'agit pas nécessairement d'une hypothèse *ad hoc*. On peut observer d'autres types de préférences temporelles qui motivent les agents à prendre des décisions distinctives à l'égard des allocations intertemporelles. Il y a, à ma connaissance, deux types de préférences temporelles dont la littérature sur le sujet fait état. Il y a les *préférences pour les délais de gratification* et les préférences pour les *séquences ascendantes*.

Une préférence pour les délais de gratification s'observe surtout chez les agents qui recherchent compulsivement des ressources en dépit du fait qu'elles leur procurent peu de gratification. Les forcenés de l'investissement et les avares surévaluent systématiquement les gains futurs par rapport aux gains immédiats. Ils sont disposés à se priver à l'extrême en vue d'une satisfaction toujours repoussée dans le temps. L'économiste Keynes fustigeait d'ailleurs ce genre d'attitude à l'égard du futur dans un propos plein d'esprit. De manière humoristique il affirmait que

> « […] *l'homme plein d'"intentionnalité", celui qui est occupé à viser des buts, est toujours en train de chercher à procurer à ses actes une immortalité illusoire et factice en projetant dans l'avenir l'intérêt qu'il peut leur porter. Il n'aime pas son chat, mais les chatons de son chat, ni même, en vérité, les chatons, mais les chatons des chatons, et ainsi de suite jusqu'à la consommation des temps dans l'univers des chats. Pour lui, de la confiture n'est pas de la confiture, à moins qu'il s'agisse d'une caisse de confiture pour demain et jamais de confiture pour aujourd'hui même. Ainsi, en rejetant toujours sa confiture loin dans l'avenir, essaie-t-il d'assurer l'immortalité à son acte confiturier.* » (1930 : 133)

Les préférences pour les délais de gratification et l'escompte hyperbolique portent sur des conséquences isolées (résultats, gratifications, moments de souffrance, gains, pertes, etc.) réalisables à deux moments du temps distincts (l'un plus près, l'autre plus loin). Loewenstein et Prelec (1993) ont mis en relief de manière expérimentale une préférence temporelle pour les séquences de conséquences qui sont présentées aux agents comme des choix explicites d'allocation intertemporelle et pas seulement pour des conséquences isolées. Les personnes préfèrent les allocations qui présentent une séquence d'amélioration aux allocations qui présentent une séquence de dépréciation de leur situation financière en dépit du fait que la séquence de dépréciation offre un rendement global légèrement plus grand. Les travailleurs préfèrent une hausse progressive de leur salaire même si la quantité globale d'argent qu'ils en obtiendraient était égale ou inférieure à une séquence décroissante. Ce type de préférence temporelle nous motive, par exemple, à choisir de recevoir les mauvaises nouvelles avant les bonnes, et pour certains, à manger leur dessert après le repas principal.

Ce qui est intéressant avec ces préférences temporelles est qu'elles poussent les agents dans des directions opposées de celles qui décrivent l'escompte hyperbolique. Les agents qui reportent systématiquement les gratifications dans le futur ne souffrent pas d'un problème de faiblesse de volonté, mais plutôt d'une forme d'*excès de volonté*. Le cas des préférences pour les séquences ascendantes pose des problèmes pour toute théorie qui expliquerait la faiblesse de volonté à l'aide de l'escompte hyperbolique. Le problème ne réside pas dans le fait qu'il existerait des motivations conflictuelles au sein des mêmes agents – ce qui est trivialement vrai –, mais dans le fait que les préférences pour les séquences ascendantes portent explicitement sur des allocations intertemporelles, alors que les préférences hyperboliques portent sur des parties de séquences, et que c'est, dans la plupart des situations, ces dernières qui l'emportent. Mais le problème n'est peut-être qu'apparent. Dans la plupart des situations, nous choisissons des items parce qu'ils exemplifient une ou deux propriétés et ignorons le reste, ou choisissons délibérément de ne pas en tenir compte lors de la computation des raisons. C'est ce qui arrive notamment lorsque nous utilisons l'heuristique de décision à *une seule raison discriminante*. Les agents akratiques peuvent choisir une allocation seulement sur la base de la seule comparaison de deux conséquences virtuellement réalisables à deux moments distincts (dans deux allocations concurrentes),

mais ils seraient néanmoins disposés à décider autrement s'ils comparaient un échantillon plus large de conséquences – comme dans les situations où on leur présente explicitement des séquences comme des parties importantes d'allocation.

Cela dit, certains théoriciens estiment que les préférences pour des séquences ascendantes ne sont pas d'authentiques préférences temporelles, car elles sont dérivées de considérations plus fondamentales. Discutant des projets de vie rationnels, Rawls adopte un point de vue semblable lorsqu'il affirme que

> « […] *toutes choses étant égales par ailleurs, nous devrions organiser les étapes les plus précoces de façon à permettre, dans celles qui leur succéderont, une vie heureuse. Il semble qu'en général il faille préférer une augmentation des attentes avec le temps. Si la valeur d'une activité est établie relativement à sa propre période, en admettant que cela soit possible, nous devrions expliquer cette préférence par les plaisirs plus grands de l'anticipation par rapport à ceux de la mémoire. Même si la somme totale de joies est la même, l'augmentation des attentes procure une satisfaction qui fait toute la différence. »*[116] (1973 : 461-462)

Si les considérations que Rawls met en relief sont celles qui motivent effectivement les agents à choisir des allocations – dans des situations où elles sont présentées explicitement comme telles – qui présentent un profil croissant, alors il ne s'agit pas de préférences temporelles comme les préférences hyperboliques et les préférences pour les délais de gratification.

10.2 La faiblesse de volonté : une conception de l'allocation intertemporelle

On est maintenant en mesure de formuler une conception de base de la faiblesse de volonté assez générale pour diagnostiquer (servir de critère pour identifier) ses épisodes, autant l'akrasie que la stricte akrasie, mais assez précise pour capturer ses traits les plus importants, et assez informative pour orienter la recherche. La conception qu'on peut formuler à partir de ce qui a été dit plus haut peut être raffinée et faire l'objet de rajouts ultérieurs. Mais elle correspond, dans une version assez détaillée, à la conception suivante:

CIT : *Un agent vit un épisode de faiblesse de volonté si et seulement s'il choisit une allocation intertemporelle en raison de préférences temporelles qui les amènent à escompter hyperboliquement le futur (à l'intérieur d'un horizon temporel catégorisé fonctionnellement).*

[116] Souligné par moi. Voir également le § 45 du même ouvrage sur l'irrationalité des préférences intertemporelles.

La CIT (pour Conception Intertemporelle de la faiblesse de volonté) comporte plusieurs éléments sur lesquels on peut s'arrêter. Le premier est que ce sont les décisions ou les choix, et non les comportements, qui sont centraux dans le phénomène (voir le chapitre 3 pour une explication). Le deuxième est que les agents choisissent des allocations intertemporelles lorsqu'ils vivent un épisode de faiblesse de volonté. Il n'est toutefois pas nécessaire de mentionner cette condition. On peut se contenter de dire que les agents choisissent en raison de préférences temporelles qui, par définition, portent sur des allocations ou des parties d'allocations (distributions plus ou moins complètes de conséquences dans le temps). Mais comme la nature de l'objet du choix importe dans une telle conception et constitue une différence majeure avec les autres conceptions de la faiblesse de volonté, il est utile de l'inclure. Le troisième élément à considérer est le rôle que jouent les préférences temporelles dans le choix d'une allocation. Lorsqu'on dit habituellement qu'un agent a choisi en raison de tel ou tel motif, on soutient par là que non seulement tel ou tel motif a fait partie de son espace motivationnel au moment où il a pris une décision, mais que c'est ce motif qui a fait la différence. Le quatrième élément exprime l'idée que les préférences temporelles doivent amener l'agent à escompter hyperboliquement le futur. Comme nous l'avons vu plus haut, il y a d'autres types de préférences temporelles (pures ou dérivées) qui n'impliquent pas un tel escompte (préférences pour les délais de gratification et préférences pour les séquences ascendantes). Il est utile de mentionner ce dernier élément même s'il n'est pas nécessaire. Les agents escomptent toujours l'avenir à l'intérieur d'un horizon temporel fonctionnellement catégorisé. Mais mentionner cet élément est pertinent dans la mesure où les agents sont susceptibles de vivre des épisodes de faiblesse de volonté lorsqu'ils s'adonnent à des activités qui s'étendent très loin dans le temps aussi bien que dans un horizon temporel de quelques secondes. Cela nous permet de traiter des cas de violation de plan financier, aussi bien que des cas de grattage compulsif des plaies, comme des épisodes de faiblesse de volonté présentant les mêmes propriétés fondamentales.

La CIT comporte évidemment des clauses implicites. Elle dit implicitement que les agents qui vivent des épisodes de faiblesse de volonté disposent d'autres options d'allocation

qu'ils seraient capables de repérer et de choisir[117]. Cela est d'ailleurs implicite à la notion même de choix : effectuer un choix implique l'existence de plusieurs alternatives.

Cela dit, la CIT est une conception, à mon avis, plus basique de la faiblesse de volonté que les conceptions qu'on trouve sur le marché philosophique (découlant de la définition standard ou d'une conception fondée sur l'irrésolution). Elle ne fait pas de différence entre les cas d'akrasie et les cas de stricte akrasie : les deux présentent les mêmes propriétés fondamentales au niveau de leurs choix d'allocation et au niveau de leurs motivations à les choisir. La formation ou non, au sein de l'agent, d'un jugement selon lequel il devrait choisir une autre allocation que celle qu'il s'apprête effectivement à choisir est une caractéristique plus accessoire que centrale. D'ailleurs, l'examen des cas de stricte akrasie qu'on peut observer dans le réel révèle les mêmes propriétés fondamentales. Nous ne décidons pas à l'encontre de ce que la tradition philosophique appelle le meilleur jugement dans n'importe quelle situation. Nous serons disposés à le faire si les choix en lice comportent des profils expérientiels spécifiques et pour lesquelles les préférences temporelles, comme l'aversion au délai de gratification ou la préférence pour le délai de frustration, feront la différence. Reste à savoir pourquoi les agents sont disposés, dans certains cas, à expliciter consciemment leur incohérence pratique (comme dans la stricte akrasie), alors qu'ils ne sont pas disposés à le faire dans d'autres cas. La manière dont les mécanismes de réduction de la dissonance cognitive s'enclenchent joue sans doute un rôle ici. Mais il s'agit d'un problème connexe et non central quand on essaie d'expliquer le phénomène de la faiblesse de volonté.

Il y a d'ailleurs un point supplémentaire sur lequel la CIT diffère d'une conception fondée sur l'irrésolution comme celle de Holton. Si l'on accepte l'idée que les comportements d'investissement compulsif correspondent à des allocations qui sont motivées par des préférences temporelles comme une préférence pour le délai de gratification, alors on doit vraisemblablement exclure *certains* cas d'irrésolution que Holton est pourtant disposé à classer dans la faiblesse de volonté. Si je suis, par exemple, résolu à reporter de manière exagérée ma consommation à un avenir lointain, mais que je faillis à la tâche en consommant maintenant une toute petite partie de ce que je m'étais pourtant engagé à

[117] Ces clauses sont également implicites dans la définition Standard.

investir, je ne souffrirai pas de faiblesse de volonté. Si, en violant ma résolution, j'opte pour une nouvelle allocation intertemporelle qui présente un profil hédonique plus adéquat et qui n'est pas motivée par une surévaluation hyperbolique des conséquences à court terme, je ne vis pas un épisode de faiblesse de volonté. Cela ne découle pas seulement de la CIT, mais d'intuitions fortes qu'on a à l'égard de ce genre de cas. Mais, la plupart du temps, lorsqu'un agent viole une résolution, il choisit une allocation inadéquate en fonction d'une préférence temporelle qui l'amène à escompter hyperboliquement le futur. C'est pourquoi on a tendance à les classer spontanément dans la catégorie des épisodes de faiblesse de volonté.

10.2.1 L'irrationalité de la faiblesse de volonté dans le cadre d'une conception intertemporelle

Est-ce qu'une conception de l'allocation intertemporelle comme la CIT permet d'expliquer pourquoi on doit estimer que la faiblesse de volonté est irrationnelle? Je pense que c'est le cas – du moins, il est beaucoup plus facile de le faire pour la CIT que dans le cadre de la conception Standard. En dévaluant le futur, les agents akratiques s'exposent plus qu'il ne le faut à l'échec. Ils n'atteignent pas ou difficilement leur objectif à long terme, et en particulier quand ces objectifs impliquent la réalisation de tâches ennuyeuses, déplaisantes ou carrément douloureuses – ce qui est déplaisant à faire, mais dont on aime le résultat. En dehors des interférences extérieures, les raisons qui motivent les agents à adopter des comportements akratiques constituent le principal écueil pour l'action planifiée. Dans un article célèbre, l'économiste R. H. Strotz (1955) avait déjà mis en relief le fait que les préférences temporelles sont une source d'inconsistances qui met en péril la réalisation de plans rationnels. Il offre en fait une démonstration technique de l'écueil que Bratman et Holton voient dans la faiblesse de volonté pour la planification rationnelle.

Or, le caractère irrationnel de la faiblesse de volonté ne se limite pas au fait qu'elle constitue un obstacle important pour la réalisation de nos plans. Faire des choix qui reflètent une dévaluation du futur altère d'une manière générale nos conditions de vie et plus spécifiquement les conditions qui nous permettent d'atteindre nos objectifs. Les travaux fondateurs de Walter Mischel, dans les années 1970, documentèrent des observations sur des enfants auxquels on offrait deux guimauves après un délai d'attente si et seulement s'ils résistaient à l'attrait d'une seule guimauve. Il s'avéra, 15 ans plus tard,

que les enfants qui s'étaient montrés plus patients avaient de meilleurs résultats scolaires, étaient plus appliqués dans la réalisation de leurs tâches scolaires, et avaient, par le fait même, beaucoup plus de chances d'entrer dans les meilleures universités (Mischel, Shoda, & Rodriguez, 1989). Joseph Heath (2009) défend l'hypothèse que ce qui expliquerait la pauvreté récalcitrante dans les pays développés (la pauvreté qui persiste en dépit d'une certaine redistribution des richesses) reposerait en partie sur l'idée que certains consommateurs sont beaucoup trop impatients d'utiliser la partie de leur revenu assignée aux dépenses discrétionnaires.

En fait, peu importe nos objectifs, nous avons besoin, pour des raisons instrumentales, de ce que Rawls (1973) a appelé des *biens premiers* comme la santé, la sécurité, les opportunités, l'intelligence, la liberté, la richesse et les pouvoirs, etc. À moins d'avoir une déficience psychologique profonde, nous préférons *ceteris paribu*s avoir davantage de ces biens que d'en avoir moins. Or, les épisodes de faiblesse de volonté nous éloignent de ces biens ou dispersent ceux que nous avons. Aussi, la faiblesse de volonté ne touche pas seulement la poursuite d'objectifs idiosyncrasiques que les agents se donnent et pour lesquels ils adoptent des plans ou forment des intentions. Elle contamine les conditions générales nécessaires à la poursuite d'objectifs quels qu'ils soient. Si notre santé, par exemple, est altérée par des comportements impatients et procrastinateurs, nous aurons beaucoup plus de mal à atteindre nos objectifs de vie, peu importe la nature de ces derniers.

Cela dit, les raisons que je viens de mentionner sont très générales, mais somme toute assez vagues. Est-ce qu'on pourrait *à l'occasion* choisir de manière rationnelle des allocations en fonction d'une préférence temporelle décrite par un escompte hyperbolique? Si cela s'avérait être le cas, alors la CIT serait incomplète dans la mesure où nous avons de bonnes raisons de croire que la faiblesse de volonté est le paradigme de l'irrationalité pratique[118]. Mais je ne pense pas que ce soit le cas. Cependant, pour montrer que la faiblesse de volonté est toujours irrationnelle, il faut, je pense, montrer pourquoi on doit estimer qu'un agent

[118] Je suis conscient que ce point ne fait cependant pas l'objet d'un consensus parmi les philosophes. Certains théoriciens comme Ogien (2003) et McIntyre (1990) estiment qu'il peut être rationnel de se comporter dans certaines circonstances de manière akratique. Mais ces auteurs supposent que la conception Standard est correcte et décrit la plupart des cas de faiblesse de volonté. En fait, des conceptions purement descriptives, comme la conception standard de la stricte akrasie ou de l'instabilité dynamique des goûts, impliquent l'existence de faiblesses de volonté rationnelles. Et il est difficile de voir comment une conception non normative pourrait rendre compte du fait que la faiblesse de volonté est toujours irrationnelle.

akratique *aurait dû* choisir une autre allocation que celle qu'il a effectivement choisie. C'est un point qui n'est cependant pas aussi évident qu'il en a l'air. La difficulté ne réside pas dans le fait que les agents akratiques peuvent à l'occasion faire des choix qui améliorent effectivement leur condition d'existence. Par exemple, je peux succomber à la tentation d'acquérir une maison luxueuse au-dessus de mes moyens parce que je veux en mettre plein la vue à mes collègues de travail, alors que cette maison s'avérera en fait être un placement financier très lucratif. Comme pour n'importe quel type de choix, on ne mesure pas son degré de rationalité strictement en fonction de la réussite ou de l'échec de ce choix. On doit examiner les raisons que l'agent avait au moment où il a pris sa décision. Si l'on estime qu'un agent s'est trompé en faisant un choix, cela n'implique pas nécessairement que son choix était irrationnel. On pourra le juger irrationnel uniquement si l'on estime qu'il avait de mauvaises raisons de faire ce choix. Par contre, cela n'implique pas que la réussite et l'échec effectif ne jouent aucun rôle dans un diagnostic de rationalité ou d'irrationalité. Avoir de bonnes raisons de prendre une décision plutôt qu'une autre est important dans la mesure où cela maximise nos chances de succès, sans être pour autant une garantie de succès (De Sousa , 2004). Aussi, lorsqu'on utilise l'expression « *aurait dû choisir* », il faut porter attention au sens qu'on lui prête. On peut l'utiliser comme jugement *post hoc* pour identifier la bonne décision qu'un agent n'a pas prise. Mais on peut l'utiliser comme jugement normatif selon lequel l'agent n'avait pas de bonnes raisons de choisir ce qu'il a effectivement choisi – en dépit du fait qu'il ait pu accidentellement réussir. Aussi, si l'on estime habituellement qu'un agent akratique n'aurait pas dû prendre la décision qu'il a prise, c'est qu'on estime qu'il n'avait pas de bonnes raisons de le faire. En ce sens, il s'agit d'un jugement normatif fondé sur des principes prudentiels plus généraux, mais aussi, en amont, sur des considérations statistiques.

10.2.1.1 La norme prudentielle fondamentale pour exclure les préférences temporelles

Les agents qui escomptent hyperboliquement le futur de manière très prononcée et dans beaucoup de champs d'activité éprouvent beaucoup de difficultés à améliorer leurs conditions de vie et sont plus prompts à dégrader celles-ci. Ils éprouvent souvent des regrets et estiment qu'ils n'auraient pas dû décider comme ils l'ont fait. Ces regrets sont justifiés dans la mesure où ils sont motivés par la conscience qu'on a enfreint une norme

prudentielle importante, à savoir que nos *décisions doivent toujours refléter une égale considération pour toutes nos tranches temporelles futures*.

Cette norme prudentielle – d'égale considération intertemporelle – n'est évidemment adéquate que dans la mesure où nos tranches temporelles sont également divisées. En effet, on ne doit pas avoir une égale considération pour des tranches qui s'étendent sur plusieurs années, d'une part, et des tranches qui ne s'étendent que sur quelques secondes, d'autre part. Cette clause est donc implicite dans l'énoncé de la norme d'égale considération intertemporelle.

Des théoriciens comme Sidgwick (1874), Rawls (1973), Elster (1979, 1986) et Broome (1991) ont insisté sur l'idée qu'un agent rationnel doit avoir une attitude de neutralité temporelle parce qu'il doit considérer – dans la mesure du possible – sa vie comme un tout et non comme un ensemble de parties disjointes. Ne pas considérer sa vie comme un tout implique que nous considérons naturellement que le « pilotage à vue » est la meilleure façon de diriger sa vie. Or, cela nous expose, par le fait même, à de plus grands écueils ou du moins à une vie peu satisfaisante.

Bien que triviale, la norme d'égale considération intertemporelle n'est pas suffisamment précise pour nous offrir des détails sur les allocations optimales. Un agent peut savoir qu'une allocation respecte la norme sans pour autant savoir s'il s'agit ou non d'une allocation optimale. En fait, comme nous l'avons vu au chapitre 6, à moins qu'ils soient confrontés à un problème décisionnel trivial – qui comporte très peu de choix, d'aspects pertinents, tous commensurables, avec peu d'incertitudes et d'impacts résiduels, etc. –, les agents rationnels ne cherchent pas à maximiser leur bien-être, mais plutôt à faire des choix satisfaisants (*satisfacing*). Avoir à trancher entre des alternatives d'allocation intertemporelle est rarement trivial. Il est très difficile d'estimer correctement, et dans le détail, les profils expérientiels afférents. La plupart du temps, nous nous basons sur des estimations sommaires obtenues en appliquant, consciemment ou non, toutes sortes de règles et de procédures décisionnelles, quelquefois sophistiquées, souvent très simples. Nous pouvons, par exemple, faire un calcul coût/bénéfice de nature strictement pécuniaire pour trancher entre l'option d'acheter une propriété ou de demeurer locataire. Nous pouvons imiter simplement notre entourage pour déterminer l'âge où il est approprié de concevoir des enfants. Aussi, je peux dans chaque cas ignorer quel est le choix d'allocation

optimale, mais tout de même avoir une égale considération pour toutes mes tranches de vie future.

En fait, cela est analogue à ce qu'on trouve dans les théories modernes de justice distributive qui impliquent un engagement envers le principe d'une égale considération de tous les citoyens, mais qui restent plus ou moins indécises sur les distributions (de ressources, de droits et d'obligations, de capabilités, etc.) les plus efficientes. La raison est simple : un principe d'égalité politique est compatible avec toutes sortes de distributions, allant des plus efficientes aux moins efficientes. L'utilité d'un principe d'égalité politique réside plutôt dans le fait qu'il joue le rôle d'un filtre pour les considérations susceptibles d'entrer en ligne de compte soit dans la réalisation effective d'une distribution ou dans le design des institutions qui devront encadrer ces distributions. Par exemple, toutes les versions acceptables du principe d'égalité politique proscrivent les critères raciaux ou sexistes comme critères de distribution.

Une norme prudentielle d'égale considération pour toutes nos tranches de vie futures joue essentiellement le même rôle. Elle ne permet pas d'identifier les allocations optimales : elle est compatible avec toutes sortes d'allocation – des meilleures aux pires. Toutefois, elle proscrit les préférences temporelles pures comme critères de choix pertinent. Aussi, s'y conformer n'est pas nécessairement un gage de succès, mais nous évite de prendre *certaines* décisions qui produisent statistiquement des profils hédoniques globalement peu avantageux.

10.2.1.2 La rationalité des préférences temporelles : deux arguments peu convaincants

Choisir une action ou une ligne de conduite simplement parce que certaines de ses conséquences positives sont plus près de nous dans le temps est irrationnel puisque cela ne nous permet pas d'améliorer notre sort aussi bien que nous le pourrions. Les préférences temporelles pures comme l'aversion au délai de gratification ou la préférence pour le délai de frustration sont soit des motifs de choix pernicieux, soit des motifs non pertinents – parce que n'ayant pas d'impact sur le bien-être global de l'agent. D'où l'évidence du principe prudentiel de l'égale considération pour tous les moments de notre vie.

Pourtant, il y a deux arguments en faveur de l'idée que les préférences qui amènent les agents à escompter hyperboliquement le futur puissent rationnellement motiver ceux-ci à prendre des décisions. Il y a l'argument du surinvestissement et l'argument ontologique.

1/ L'argument du surinvestissement est une sorte de preuve par l'absurde. Nous devons faire des choix en fonction de préférences temporelles pour le court terme parce que sans cela nous repousserons systématiquement à plus tard les activités plaisantes et la consommation de biens. Il s'ensuit que nous devons faire des choix motivés par des préférences temporelles. On trouve des versions plus ou moins élaborées de cet argument dans les justifications que certaines personnes font de leurs mauvaises habitudes de consommation. Sans faire preuve de mauvaise foi, certains fumeurs assument leur dépendance et expriment leurs convictions mollement en disant : « *il faut profiter des choses pendant qu'elles passent… parce qu'on n'a pas toute la vie!* ».

Il y a deux critiques qu'on peut faire à l'argument (*1A et 1B*) :

(*1A*) La première critique qu'on peut faire est qu'il présente un faux dilemme. On ne doit pas choisir en fait entre le court terme et le long terme, mais entre le court terme, le long terme et toute une série d'allocations qui distribuent de manière plus uniforme les conséquences sur le long terme *et* le court terme.

(*1B*) La seconde critique est qu'il pourrait au mieux-être recevable dans un contexte où chacun de nous est immortel. Pour un organisme immortel, il serait peut-être rationnel pour lui de reporter dans le temps ses périodes de consommation de biens si cela lui permettait, grâce à un certain investissement, d'obtenir un bien futur plus grand. De la même manière, il serait rationnel pour lui de devancer la réalisation de maux, si cela lui permettait d'éviter un plus grand mal dans l'avenir. Le problème est que nous ne sommes pas immortels, donc nous n'avons pas de raisons de privilégier systématiquement l'avenir.

2/ L'argument ontologique est plus sophistiqué, voire exotique. Il a été formulé par le philosophe Derek Parfit (1984) et figure dans son ouvrage majeur, ***Reasons and Persons***. Il consiste en gros à dire que, comme notre concept de personne a une dimension fictionnelle importante, il n'y a pas de raisons de ne pas dévaluer l'avenir. Parfit se sert d'une expérience de pensée inspirée de la science-fiction pour appuyer l'hypothèse que le concept de personne est plus une chimère verbale qu'une réalité, bien qu'il soit implémenté dans la

culture occidentale. Si, sur Terre, on met une personne dans un téléporteur qui doit détruire la structure organique de celle-ci pour en tirer l'information nécessaire (sorte de numérisation agressive!) pour reconstruire ensuite cette personne sur Mars, est-ce que la personne qui en résulte est la même que celle qui est entrée dans le téléporteur sur Terre? On est tenté de répondre par l'affirmative d'autant plus que la personne sur Mars a les mêmes traits psychologiques que celle sur Terre : elles partagent toutes les deux les mêmes souvenirs, les mêmes croyances et les mêmes préférences, etc. Mais qu'arriverait-il si, à partir des informations obtenues par le premier téléporteur, on créait deux personnes? Qui serait la première personne? Pour Parfit, il s'agit d'un problème insoluble qui devrait faire suffisamment douter du concept d'identité personnelle pour qu'on le rejette. Parfit propose de le remplacer par celui de connexion psychologique. Or, pour le problème qui nous intéresse, la notion de proximité psychologique pourrait expliquer pourquoi nous dévaluons le futur, et dans une certaine mesure, justifier cette attitude prospective. Pour Parfit, nos relations avec nos alter ego sont essentiellement les mêmes qu'avec les autres personnes. Plus nous sommes proches d'une personne, plus nous sommes disposés à nous en soucier, mais plus elle est éloignée, moins nous nous en soucions. De manière analogue, plus la connexion est grande avec nos alter ego futurs, plus nous sommes disposés à en tenir compte lorsque vient le temps de faire des choix. Inversement, moins elle est grande et moins nous nous en soucions. Or, comme nos alter ego situés très loin dans le temps ont une moins grande connexion avec nous que ceux qui vivront dans un avenir plus proche, il s'ensuit que nous nous soucierons davantage de ces derniers. Aussi, on ne doit pas dire qu'il s'agit d'une faute prudentielle, mais seulement d'une attitude aussi justifiable qu'un traitement différencié de ses proches par rapport aux étrangers. D'ailleurs, Parfit ne dit pas qu'on ne doit jamais se soucier du futur, mais seulement qu'on doit s'en soucier en fonction d'une connexion avec son moi présent.

Il y a au moins trois critiques qu'on peut adresser à cet argument.

(*2A*) Le premier concerne la portée normative de l'absence d'identité personnelle diachronique. Peut-être que le fait que nos alter ego soient éloignés dans le temps nous permet d'en tenir moins compte lorsque vient le temps de prendre des décisions qui auront des conséquences à long terme. Mais cela n'implique pas qu'on *doive* moins en tenir compte. Par analogie, des grands-parents ne doivent pas moins tenir compte de leurs petits-

enfants que de leurs enfants, même si ces derniers sont psychologiquement plus liés à eux. Si c'était le cas, alors on devrait considérer les grands-parents « gâteaux » comme des personnes injustes, ce qui est plutôt contre-intuitif.

(*2B*) En fait, Parfit n'offre pas une justification rationnelle de la dévaluation du futur décrit comme un escompte hyperbolique. Il identifie plutôt un critère fondamental *atemporel* et affirme clairement dans certains passages que l'escompte d'un agent qui dévalue le futur n'est pas motivé par une préférence temporelle pure :

> « *My connectedness to my future may correspond to the degree of connectedness between me now and myself in the future* […] *since connectedness is nearly always weaker over long periods, I can rationally care less about my further future. This claim defends a new kind of discount rate. This is a discount rate, not with respect to time itself, but with respect to [connectedness]…*» (1984 : 313-314)

Même si nous acceptions la thèse de la connexion psychologique, que nous rejetions la fiction de l'identité diachronique de la personne et que nous justifiions l'idée d'une dévaluation des conséquences éloignées dans le temps, il resterait à justifier l'idée que cette dévaluation doit avoir la forme d'un escompte hyperbolique. Or, la marche est haute. Parfit ne fournit en fait qu'un paramètre externe analogue à l'incertitude et à l'instabilité anticipée des goûts. Comme il ne nous montre pas en quoi il serait rationnel d'avoir des préférences temporelles *pures* comme celles qui expliqueraient les choix d'allocations typiques des agents akratiques, il ne fournit pas une justification rationnelle de la faiblesse de volonté.

(*2C*) La seconde critique est de nature empirique. Le théoricien de la décision Shane Frederick (2003) a voulu tester si la proximité psychologique que les agents disent avoir avec leur moi antérieur ou ultérieur expliquerait et justifierait une dévaluation du futur comme le suppose Parfit. Dans quelle mesure, demande Frederick, les agents ne sont pas rationnellement tenus de se soucier de leur alter ego futur *au-delà* de ce qui est requis par leur connexion psychologique? Frederick a fait passer des tests d'estimation prospective et rétrospective du degré de similarité psychologique que des personnes pensent avoir avec elles-mêmes à divers moments du futur et du passé. Il a tenté d'observer sans succès une corrélation des résultats obtenus avec la disposition qu'a chacune de ces personnes à escompter le futur de manière hyperbolique (disposition qu'il a mesuré avec des questionnaires portant sur des alternatives de « gain versus gain » typique pour établir en laboratoire le taux d'escompte d'un agent). Il en est aussi ressorti le fait intéressant que les

personnes ne regrettent pas moins des actes passés même si ceux-ci ont été commis par des alter ego psychologiquement très éloignés d'eux (ex. : le regret d'avoir volé un jouet alors qu'on était enfant et très immature). Par contre, les résultats de l'enquête de Frederick n'impliquent pas que les agents ne dévaluent pas le futur en partie en fonction de facteurs de connexion psychologique, mais seulement qu'il ne le dévalue pas de manière hyperbolique.

10.3 Les préférences temporelles pures : conséquences et hypothèses concurrentes

L'hypothèse de l'existence de préférences temporelles pures a donné lieu à une abondante littérature et à une série d'expérimentations en milieu contrôlé (Elster & Loewenstein, 1992 ; Loewenstein, Read, & Baumeister, 2003). Le nombre de problèmes et de questions qu'elle soulève, ainsi que ses conséquences pour toute théorie empirique de la décision, est énorme. Aussi, je ne compte ici mentionner qu'une conséquence importante, à mon avis, pour le traitement philosophique du problème de la faiblesse de volonté. Je compte également me demander si les supposées préférences temporelles sont aussi pures (non dérivées) ou aussi situées à la base de l'échafaudage motivationnel des agents que certains auteurs (Ainslie & Haslam, 1992) le prétendent. Je mentionnerai seulement quelques critiques, parmi les plus intéressantes et les plus significatives pour le débat philosophique. J'aborderai deux hypothèses concurrentes pour expliquer l'escompte hyperbolique qui sont particulièrement intéressantes. La seconde des deux l'est davantage non seulement parce qu'elle a acquis une certaine robustesse depuis les dernières années, mais aussi parce qu'elle est susceptible de fournir un appui important aux théories d'inspiration platonicienne qui font une large part aux considérations de nature cognitive dans l'explication des décisions akratiques.

10.3.1 Expliquer la dévaluation du futur ou le contrôle de soi?

Une des conséquences qui découlent de l'hypothèse qu'il y a des préférences temporelles pures, et que c'est effectivement ce qui motive les agents à prendre des décisions akratiques, est que ce qui nécessite une explication est moins le phénomène de la faiblesse de volonté que le phénomène du contrôle de soi. Ainslie et Haslam estiment que les

approches cognitives de l'escompte hyperbolique sont insatisfaisantes parce que les informations qu'un agent acquiert sur les illusions motivationnelles produites par un effet de rapprochement temporel n'affectent pas sa motivation. Par contraste, le fait de connaître la nature des illusions d'optique change notre approche de ces problèmes. Un œil éduqué arrive à estimer correctement l'amplitude et les proportions des objets situés à différentes distances. Et quand nos estimations comportent une marge d'erreur trop grande, nous utilisons volontiers des instruments de mesure pour les éviter. Aussi, nous ajustons nos évaluations en fonction des transformations dans notre champ visuel et de toutes sortes d'informations auxiliaires sans ressentir une forme ou une autre de conflit interne. Pour Ainslie et Haslam,

> « […] *this is the heart of the temporary preference hypothesis. The original evaluation of delayed goods takes place in the same way as the perception of other magnitudes, but a person cannot learn to correct it as well* […].*A larger image on the retina does not of itself motivate a person one way or another and, thus, does not resist transformation by abstraction. Satisfaction, on the other hand, is the fundamental selective force of choice, and however the person perceives or categorizes it with his telescopic faculty, he is still acted upon by its direct influence. That is to say, there is a raw process of reward that constitutes the active determinant of value. While value can be perceived abstractly, it does not motivate differently because of this abstraction. Abstraction occurs downstream, as it were, from where motivation occurs.* » (1992 : 73)

Le processus *brut* dont parlent Ainslie et Haslam reflèterait la dynamique motivationnelle de base de bon nombre d'organismes et ne serait pas paramétrable *en amont* par des facteurs cognitifs d'évaluation. La saillance motivationnelle que les conséquences acquièrent à mesure qu'elles se rapprochent temporellement reposerait sur des considérations désidératives aussi brutes que notre goût inné pour des aliments sucrés et riches en matières grasses, nos préférences à la base de notre orientation sexuelle, notre aversion à la douleur, notre désir de reconnaissance sociale, etc. Si les préférences temporelles font effectivement partie du soubassement motivationnel des agents, alors notre attention théorique doit porter sur les mécanismes de contrôle de soi et non sur la faiblesse de volonté. Parce que les agents respectent parfois leurs résolutions, parfois les violent, il semble arbitraire de dire que c'est le premier phénomène, et non le dernier, qui requiert une explication spéciale. Mais l'histoire de la physique nous offre une raison de penser que cela est une bonne attitude épistémique. Se référant à la 1re loi du mouvement dans la physique newtonienne, Ainslie et Haslam affirment que

> « […] *when Aristotelian philosophers looked for propellants rather than retardants, they actually found none and had to account for momentum with ad hoc constructs that lacked predictive power, such as streams of air doubling back from the front of a thrown stone to push it from behind. When we have looked for factors that change preference rather than maintain them, we seem to have been equally misled; existing theories of "irrational" or ambivalent behavior are a hodgepodge. Given experimental evidence that preference intrinsically tends to change as a function of time, it makes sense to look instead for what factors produce constancy*. » (*ibid*.: 73-74)

Ce renversement d'attitude vis-à-vis la faiblesse de volonté nous amène à reconsidérer certains problèmes, notamment celui du conflit interne que vivent les agents akratiques. On devrait considérer des conflits de ce genre comme relativement périphériques et pas nécessairement présents lors d'épisodes de faiblesse de volonté. Les théories classiques ne peuvent expliquer pourquoi les agents ne vivent pas de conflit motivationnel lorsque les conséquences des choix présents ne se réaliseront que dans un futur lointain. C'est la présence occasionnelle des conflits qui nécessite une explication spéciale et pas le fait que les agents aient des préférences temporelles qui les amènent à escompter hyperboliquement le futur. Une hypothèse, par exemple, serait que l'adhésion plus ou moins consciente au principe d'égale considération intertemporelle est un facteur de conflit interne (ou même une simple adhésion à l'idée que nos choix doivent être cohérents) lorsque les conséquences des alternatives d'allocation se rapprochent suffisamment de nous. Le conflit lui-même est insoluble parce que chacune des alternatives d'allocation est dominante à différents moments. L'existence de préférences temporelles pures au sein de l'espace motivationnel des agents n'implique pas qu'ils succombent toujours à la tentation ou que leurs conflits intrasubjectifs sont invariablement résolus à la faveur de ces préférences. C'est plutôt qu'en l'absence de motivations contraires suffisamment fortes, les agents choisiront en raison d'une préférence temporelle[119].

[119] Évidemment, poser l'hypothèse de préférences temporelles pures n'implique pas que celles-ci n'ont pas de paramètres plus fondamentaux. Baumeister et coll. (Baumeister, Heatherton & Tice, 1994) ont proposé les paramètres d'*impulsivité* (le degré auquel un agent choisit de manière spontanée), de *compulsivité* (le degré auquel un agent fait des plans, prend des résolutions et les exécute) et d'*inhibition* (la capacité d'inhiber les réponses automatiques à des motivations viscérales comme les appétits et les émotions qui déclenchent de telles réponses) pour expliquer le poids des préférences temporelles. Il ne s'agit toutefois pas de motifs à proprement parler – ce ne sont pas des raisons de choisir –, mais plutôt de facteurs importants affectant certains de nos motifs de choix – les préférences temporelles en premier lieu.

10.3.2 Escompte hyperbolique sans préférences temporelles pures

L'hypothèse qu'il y ait des préférences temporelles pures qui amènent les agents à escompter hyperboliquement le futur ne fait pas l'unanimité. Il y a plusieurs hypothèses concurrentes et documentées à divers degrés dans la littérature. Je ne compte pas les présenter et les examiner toutes ici – étant donné l'espace que cela prendrait. Mais il y a deux hypothèses qui présentent un certain attrait philosophique et une certaine robustesse empirique. La première est une hypothèse proposée initialement par Rachlin (1992, 2006) et Raineri (1992) selon laquelle l'escompte du futur n'est pas la seule fonction hyperbolique de l'esprit et qu'elle ne découle donc pas nécessairement de préférences temporelles *pures*. La seconde fut proposée par Liberman et Trope (Liberman & Trope, 1992 ; Liberman, Trope, Mccrae & Sherman, 2002) et repose sur l'idée que les agents escomptent le futur en fonction d'un niveau de construction cognitive de leurs détails. Examinons ces hypothèses tour à tour.

1/ La fonction hyperbolique répandue des évaluations et la nécessité des signaux hétérogènes. Dans un article important et critique de la position d'Ainslie et Haslam, Rachlin et Raineri montrent que la fonction hyperbolique qui décrit l'escompte du futur décrit également un tas d'autres processus cognitifs qui n'ont rien à voir avec les préférences temporelles. Ils mettent même en relief certains processus purement physiques qui produisent une évaluation hyperbolique. Par exemple, que l'évaluation, à l'aide d'un détecteur de l'intensité d'un son émis par une source, décroît en raison de l'inverse du carré de la distance qui sépare le détecteur et la source. On observe, dans ce cas, un renversement d'ordre d'intensité des sources à une certaine distance si les sources sont elles-mêmes situées à des distances différentes du détecteur. Hypothétiquement parlant, l'accroissement de l'intensité mesurée devrait se conformer à une fonction exponentielle (taux fixe). Or, étrangement, la fonction d'accroissement réellement observée est plutôt hyperbolique. La fonction d'accroissement hyperbolique s'observe pour une source d'énergie en général (son, lumière, etc.). La même chose vaut pour l'évaluation dans le champ visuel de la grandeur relative des objets – laquelle est déterminée par une sorte de fonction hyperbolique géométrique[120]. Dans la sphère psychologique, la mémoire de rappel offre

[120] Voir, par exemple, l'illusion d'optique dans la section sur la caractérisation de l'escompte hyperbolique.

également un exemple de fonction hyperbolique. La saillance d'un événement passé dans la mémoire décroît moins lentement au fur et à mesure que le moment d'apprentissage et le moment du rappel croissent (*Jost's fist law*) – ce qui crée une forme de myopie rétrospective. Mais on observe une fonction hyperbolique à l'œuvre dans la rapide dépréciation des automobiles neuves, ainsi que dans la disposition généreuse décroissante à donner à autrui à mesure que la distance sociale qui le sépare de soi s'agrandit (Rachlin, 2006). Évidemment, les mécanismes fondamentaux qui sous-tendent la fonction hyperbolique dans chacun de ces cas de figure ne sont pas de même nature. Ce qui explique la dépréciation rapide des automobiles neuves (asymétrie d'informations dans le marché de l'usager, bassin plus grand d'acheteurs pour les vieilles bagnoles, etc.) n'explique pas la différence de mesure des sources d'énergie situées à des distances différentes. Aussi, ce qui expliquerait hypothétiquement l'escompte du futur, à savoir des préférences temporelles pures, n'est évidemment pas requis pour expliquer les autres phénomènes hyperboliques de mesure.

Par contre, Rachlin et Raineri mettent en relief le fait intéressant que la fonction hyperbolique des phénomènes mesurés disparaît lorsque plusieurs sources d'informations ou de signaux sont mises à contribution. Par exemple, imaginez que vous soyez plongé dans une chambre noire et que vous ayez à évaluer la distance relative de deux poteaux lumineux qui se rapprochent de vous. Vous devez le faire avec un seul œil ouvert, vous ne pouvez pas bouger ni voir quoi que ce soit d'autre dans la chambre, et vous ne connaissez pas non plus la position initiale des poteaux. Vous allez juger la grandeur relative des poteaux strictement en fonction de la dimension de l'image rétinienne qu'ils produisent. Or, cette image croît de manière hyperbolique, ce qui vous amènera donc à considérer à tort un poteau plus grand que l'autre parce qu'il entrera à l'intérieur d'une certaine distance de vous. Mais dès que les lumières sont allumées, le second œil ouvert, et qu'il vous est permis de bouger, vous n'allez plus vous laisser berner. La computation d'une multitude de signaux pertinents passifs ou que vous sollicitez activement dans l'environnement (ex. : en bougeant) vous permettra d'obtenir une évaluation plus linéaire que si vous ne vous serviez que du seul signal de la grandeur de l'image rétinienne dans un seul œil.

Dans le cas de la motivation, ces auteurs estiment que c'est souvent le manque de signaux ou d'échafaudages actifs d'incitation dans l'environnement qui expliquerait les cas de

faiblesse de volonté. Par exemple, la célèbre expérience de Mischel sur la résistance des enfants à l'attrait d'une guimauve montre ceux-ci dans une situation où ils doivent attendre pendant plusieurs minutes sans rien faire d'autre, et sans qu'aucune personne en autorité ne leur dise qu'il est préférable d'attendre pour obtenir deux guimauves plutôt qu'une. Dans cette situation, les enfants ne disposent de rien d'autre que d'un plaisir d'anticipation faible – par comparaison avec l'excitation que produit en eux la vue d'une guimauve présente – pour résister à la tentation.

La plupart du temps, nous ne nous laissons pas berner par les distances relatives différentes des objets par rapport à nous lorsque vient le temps d'évaluer leur grandeur. De même, pour Rachlin et Raineri, les épisodes de faiblesse de volonté sont relativement peu fréquents dans la journée d'un agent « normal ». Les agents utilisent une panoplie de stratégies de contrôle de soi et d'évaluation du futur qui rendent l'exécution de leurs plans plus que probable. Ils évitent les sources de stimuli délétères (ex. : prendre un autre chemin que celui qui implique de passer devant un *fast-food*), modifient leur environnement immédiat pour stabiliser leur attention sur les tâches qu'ils doivent accomplir (ex. : vendre son téléviseur ou s'enfermer dans un cubicule à la bibliothèque), font des paris privés (ex. : si je tonds le gazon aujourd'hui, je m'autorise exceptionnellement à une glace au chocolat) ou publics (ex. : dire à tous ses amis qu'on a arrêté de fumer), se donnent des sous objectifs (ex. : aujourd'hui, je dois écrire deux pages de ma thèse), « redécrivent » des choix pour stabiliser leurs préférences (ex. : ne pas se dire qu'on a acheté un nouvel ordinateur, mais dilapidé ses économies) ou associent des activités hédoniques de qualités opposées pour raffermir leur motivation (ex. : nettoyer sa salle de bain en écoutant de la musique), etc.

Aussi, nous n'avons pas vraiment besoin de postuler l'existence de préférences temporelles pures pour expliquer les épisodes de faiblesse de volonté. Il est théoriquement plus prudent de dire que les agents escompteront hyperboliquement le futur s'ils sont « contraints » de ne se rabattre que sur la computation de certains signaux – de manière analogue à l'évaluation des grandeurs.

2/ La théorie des niveaux de construction mentale des événements. Il est possible d'expliquer pourquoi les agents escomptent hyperboliquement le futur sans postuler l'existence de considérations désidératives spéciales au sein de son espace motivationnel. Le paramètre cognitif des décisions peut être suffisamment complexe et riche pour nous

fournir les clefs nécessaires pour expliquer la faiblesse de volonté. La théorie des niveaux de construction mentale des événements fournit, aux dires de leurs promoteurs, ce genre de clefs. Les hypothèses qui forment le cœur de cette théorie sont que (*a*) le niveau construction mentale des événements virtuellement réalisables par les agents détermine en partie leur motivation à les réaliser ; et que (*b*) la distance temporelle d'un événement détermine en partie son niveau de construction mentale.

Par niveau de construction mentale, Liberman et Trope entendent la manière dont on catégorise les événements ou les objets. On classifie spontanément les choses à l'intérieur de catégories en dépit du fait qu'elles ne sont pas identiques. Aussi, catégoriser consiste à ignorer implicitement des caractéristiques uniques, mais à prendre une décision épistémique sur les caractéristiques communes plus centrales et celles qui sont plus périphériques. Catégoriser un objet comme une chaise consiste à mettre l'emphase sur des caractéristiques stéréotypiques de la plupart des chaises, comme celle de servir de « repose fesses » et celle d'être une pièce d'ameublement. Ces caractéristiques sont centrales lorsqu'on utilise la catégorie « chaise », alors que la couleur ou la matière restent périphériques. En revanche, lorsqu'on catégorise le même objet comme une pièce de bois, alors la fonction de repose-fesses et le fait d'être une pièce de mobilier deviennent périphériques, et la matière constitutive devient plus centrale. Ce qui est central et périphérique dans les catégories est sujet à des variations idiosyncrasiques. Le plaisir peut être une caractéristique centrale ou périphérique des rapports sexuels dépendamment de son *background* culturel ou émotionnel.

Beaucoup de catégories sont organisées autour de buts. Pour ces catégories, les caractéristiques seront centrales ou périphériques relativement à ces buts. Par exemple, si nous faisons un régime minceur, les calories deviendront une caractéristique centrale de la catégorie des aliments, tandis que leur bon goût, leur texture, leur degré de craquant, et leur caractère élaboré seront plus périphériques. Comme tous les autres types de catégories, les catégories orientées-vers-des-buts sont « subsumables » sous d'autres catégories orientées-vers-des-buts plus générales, de la même manière que les buts peuvent être inclus dans des hiérarchies de buts plus généraux. À l'intérieur de telles hiérarchies, les choix ou activités sont intégrés verticalement de manière ascendante (*at the superordinate level*) par la réponse qu'elle donne à la question du « *pourquoi* doit-il être réalisé? », ou de manière

descendante (*at the subordinate level*) par la réponse qu'elle donne à la question du « *comment* doit-il être réalisé? ». Liberman et Trope distinguent les niveaux d'intégration afférents des caractéristiques orientées-vers-des-buts comme les caractéristiques construites au *bas niveau* et des caractéristiques construites au *haut niveau*. Les caractéristiques de haut niveau et de bas niveau ne se distinguent pas seulement à partir de leur mode d'intégration dans une hiérarchie des buts, mais aussi à partir du changement de comportement que leur altération est susceptible d'engendrer chez l'agent.

Habituellement, modifier des caractéristiques de haut niveau produit des changements plus substantiels que l'altération de caractéristiques de bas niveau. Par exemple, changer le sujet d'une conférence annoncée est un changement plus significatif que le changement de son heure de programmation[121]. Si vous demandez à quelqu'un pourquoi il souhaite assister à telle ou telle conférence, il ne dira pas que c'est parce qu'elle a lieu à telle heure, mais parce qu'elle porte sur tel sujet. L'heure de présentation est une caractéristique périphérique des conférences, et nous aide plus à répondre à la question de savoir comment pourrais-je y assister (ex. : je devrais prendre un taxi et non les transports publics parce que je veux être sûr d'arriver à l'heure pour le début de la conférence), que de savoir pourquoi je devrais y assister.

Les auteurs distinguent également les constructions de niveaux différents par leur degré d'abstraction. Les constructions comme « élargir son esprit » sont typiquement plus abstraites que des constructions comme « lire un livre » ou plus concrètes comme « tourner des pages », même si elles s'appliquent aux mêmes activités.

(*a*) Une des hypothèses que Liberman et Trope défendent est que le niveau des constructions catégorielles des choix et des activités a une incidence importante sur leur attrait motivationnel. Cette incidence ne porte pas tant sur le degré de désirabilité des choix et des activités (bien qu'elle le fasse en partie) que sur leur qualité hédonique. Les constructions de haut niveau confèrent aux choix et aux activités un caractère plaisant – ou du moins attirant –, alors que les constructions de bas niveau leur confèrent un caractère déplaisant – ou du moins ennuyeux. Acquérir une grande maîtrise de son instrument de musique favori est un choix plus attirant que de faire des gammes avec un métronome

[121] Exemple de Liberman et Trope.

plusieurs heures par jour, même s'il s'agit foncièrement de la même activité. Contrairement à la question du pourquoi, la question du comment produit typiquement une déflation dans l'attrait motivationnel d'un choix ou d'une activité. Nous sommes souvent enthousiastes à l'idée de faire telle ou telle activité, mais cet enthousiasme fond, à notre grand étonnement, comme neige au soleil à partir du moment où nous commençons à considérer les moyens concrets qu'il faut mettre en œuvre pour l'accomplir. Aussi, la catégorisation avec le « pourquoi » en trame de fond a un impact motivationnel plus positif qu'une catégorisation avec le « comment », qui a un effet contrastant nettement négatif.

(*b*) La seconde hypothèse de Liberman et Trope est que la distance temporelle d'un choix ou d'une activité affecte notre catégorisation de celles-ci. Nous utilisons spontanément des catégories de haut niveau pour des actions plus éloignées dans le futur et des catégories de bas niveau pour les actions envisagées dans un avenir rapproché. Par exemple, on va davantage envisager un déménagement comme le commencement d'une nouvelle vie s'il a lieu dans un futur lointain, mais on l'envisage comme le transport de boîtes s'il a lieu demain matin. De même, des sujets auxquels on demande d'imaginer qu'ils ferment à clef la porte de leur domicile demain ne décriront pas de la même façon l'activité si on leur demande d'imaginer le faire dans une année. Dans le premier cas, ils utiliseront une catégorie plus concrète comme « tourner la clef dans la serrure », alors qu'ils diront plus spontanément qu'ils sécurisent leur domicile dans le second cas.

On peut se demander pourquoi on associe la construction de niveaux catégoriels à l'escompte du futur. Une hypothèse plausible serait de dire que les informations que les agents ont à l'égard du futur sont d'autant moins fiables que ce futur est lointain. Mais, en fait, les agents continuent de former des niveaux de construction catégorielle différents même quand les détails des événements lointains sont connus ou faciles à obtenir. Une autre hypothèse consiste à dire que plus les événements sont temporellement lointains, plus il est facile de revenir sur sa décision, ce qui rend peut-être inutile la planification des moyens. Mais, comme Liberman et Trope l'affirment,

> « [la théorie des niveaux de construction] *propose that individuals continue to use high-level construals for distant future events and low-level construals for near future events even when the above mentioned reasons do not exist – that is, when the information about near future and distant future events is the same and an irreversible decision has to be made at both point of time. Our research suggests that*

> *temporal construal is an overgeneralised tendency that is applied in situation in which it is neither appropriate nor necessary.* » (1992 : 249)

Les auteurs soutiennent ici clairement qu'il s'agit d'un phénomène robuste. Il est assez robuste à leur avis pour expliquer des cas typiques de renversement de préférences, d'irrésolution et de violation de plan qui découlent d'un escompte hyperbolique du futur – mais sans faire appel à des préférences temporelles. La conjonction de (*a*) et (*b*) permet en effet d'expliquer le changement intertemporel des valeurs attribuées aux conséquences futures :

> « […] *it predicts that the value associated with low-level construals should be more prominent in a subjective evaluation of near future events, whereas the value associated with high-level construals should be more prominent in evaluating distant future events.* » (*ibid.*: 250)

On peut avoir de bonnes raisons d'adopter des niveaux de construction différents pour les événements lointains et rapprochés – comme les deux mentionnées plus haut. Mais il ne fait pas de doute que cela constitue un obstacle de taille pour l'action planifiée. Et qui plus est, c'est un obstacle qu'on ne remarque pas facilement. Toutefois, il n'est pas certain que la théorie des niveaux de construction soit plus appropriée pour expliquer la faiblesse de volonté que l'hypothèse des préférences temporelles pures. Elle pourrait très bien expliquer pourquoi nous faisons si volontiers des plans irrationnels pour l'avenir lointain. C'est un point qui est d'ailleurs mis en relief par le psychologue Daniel Gilbert (2006). Comme un champ de maïs observé du ciel, beaucoup d'événements lointains présentent « une douceur » qu'ils perdent lorsqu'on les contemple de plus près[122]. Cela pousse des personnes à s'engager dans des activités qu'ils sont susceptibles de regretter par la suite. Je peux offrir des services de gardiennage à une tante qui part une semaine pour l'étranger parce que je me représente l'événement comme une occasion de me rapprocher de mes neveux et nièces. Mais je peux m'en mordre les doigts la veille parce que je traite mentalement des informations pour déterminer comment je devrais passer la semaine avec eux, et en particulier comment je devrais passer la journée de demain avec eux. Faire du gardiennage dans un mois est une manifestation d'amour, en faire demain, c'est préparer le petit-déjeuner. Aussi, la théorie des niveaux de construction semble plus appropriée pour une théorie plus générale du *mal-vouloir* (se tromper à vouloir quelque chose) que de la

[122] Métaphore qu'il emprunte à Newby-Clark et Ross (2003).

faiblesse de volonté. Bien que la faiblesse de volonté devrait idéalement figurer dans un chapitre d'une théorie générale du mal-vouloir, l'absence d'arrimage avec ce qui arrive typiquement aux agents akratiques constitue un défaut. L'hypothèse des préférences temporelles semble à cet égard plus appropriée.

10.4 Conclusion

Les attitudes orientées vers l'avenir, et non le souci de cohérence avec des jugements ou des décisions passés, sont la clef pour expliquer les épisodes de faiblesse de volonté, mais aussi les décrire correctement. Le schème de la théorie du choix intertemporel fournit les meilleurs outils conceptuels pour aborder le phénomène. Du point de vue empirique, on constate qu'il n'y a pas de différences significatives entre les *choix* strictement akratiques et ceux relevant de la faiblesse de volonté diachronique. Les deux impliquent des distributions de conséquences présentant des profils hédoniques analogues. La différence ne réside pas, à proprement parler, dans les propriétés de ces choix, mais dans le fait qu'il y a ou non présence d'un conflit interne qui tire l'agent dans des directions opposées.

Les motifs qui expliquent pourquoi les agents optent à l'occasion pour des choix désavantageux (ou pas aussi avantageux qu'ils pourraient l'être) *typiques de la faiblesse de volonté* résident dans le fait qu'ils surévaluent les conséquences à court terme par rapport à celles réalisables dans un horizon temporel plus lointain. Cette dévaluation couplée avec la structure temporelle des allocations disponibles explique l'attrait des choix akratiques.

La nature de ces motifs qui sous-tendent la dévaluation du futur ne fait cependant pas l'unanimité parmi les chercheurs. Bien qu'on sache qu'ils prennent synthétiquement la forme d'un escompte hyperbolique dans le cas de la faiblesse de volonté, il n'est pas clair que le motif responsable des épisodes de faiblesse de volonté soit une préférence temporelle pure comme l'aversion pour le délai de récompense ou la préférence pour le délai de frustration. La présence d'une préférence temporelle au sein de l'espace motivationnel de l'agent est sans doute l'hypothèse la plus simple et la plus naturelle. Or, il est difficile de dire avec certitude s'il s'agit vraiment d'une préférence non dérivée de considérations cognitives plus fondamentales concernant la perception de l'avenir et la capacité d'en anticiper les détails. D'ailleurs, comme l'escompte hyperbolique ressemble à d'autres fonctions hyperboliques– en particulier celles qui concernent les perceptions

spatiale et auditive – sur la base desquelles l'esprit fonde certaines évaluations en l'absence d'indicateurs hétérogènes variés, il est difficile de dire si les supposées préférences pures, constitutives de notre espace motivationnel, ne sont pas en fait formées seulement en l'absence d'incitations variées.

La théorie du choix intertemporel soulève beaucoup de questions à l'égard de certains aspects de la faiblesse de volonté, mais aussi beaucoup de questions au sujet des mécanismes de contrôle de soi. Adhérer au principe d'égale considération pour toutes nos phases temporelles futures est rationnellement justifiable à la lumière des meilleures raisons, mais n'est qu'un élément parmi d'autres, susceptibles d'enclencher de tels mécanismes. L'adhésion à ce principe est au cœur de bon nombre (mais pas de tous) de conflits internes typiques de la stricte akrasie, mais l'enclenchement des mécanismes de contrôle de soi repose sur toute une série de facteurs qui nécessite autant d'explications que le phénomène de la faiblesse de volonté lui-même.

11
Conclusion

L'instabilité de l'état social vient à favoriser l'instabilité naturelle des désirs. Au milieu de ces fluctuations perpétuelles du sort, le présent grandit; il cache l'avenir qui s'efface, et les hommes ne veulent songer qu'au lendemain.

– Tocqueville, De la démocratie en Amérique

11.1 Notes sur les stratégies individuelles et collectives de contrôle de soi

La description et l'analyse des choix akratiques sont des tâches importantes en philosophie, en psychologie et dans les sciences sociales parce que les mener à bien nous permettrait de comprendre et d'expliquer pourquoi les agents humains (et non humains) agissent souvent à l'encontre de leurs propres intérêts ou échouent à améliorer suffisamment leur condition d'existence en dépit du fait qu'aucun obstacle ne leur barre la route. La faiblesse de volonté est en ce sens le paradigme de l'irrationalité pratique. Mais elle joue également un rôle important dans l'explication de certains comportements non coopératifs et immoraux.

Les analyses purement conceptuelles, comme on en trouve beaucoup dans la littérature philosophique, ont à cet égard une utilité limitée. De même, les disputes entourant l'existence présumée de la stricte akrasie ont tendance à faire dévier l'enquête sur autre chose que les choix et les motivations effectives des agents akratiques et à oblitérer leurs propriétés les plus significatives. Or, cela ne nous met pas en bonne position pour élaborer des stratégies individuelles de contrôle de soi et encore moins faire des recommandations pour des politiques publiques.

En revanche, si l'on accepte l'hypothèse minimale que les agents akratiques escomptent hyperboliquement le futur, on sera en meilleure position pour évaluer ou élaborer des stratégies personnelles et des dispositifs institutionnels commandés par l'intérêt bien pesé. Les personnes peuvent bien adhérer au principe d'égale considération pour chacune de leur phase temporelle, ils doivent néanmoins adopter des stratégies ou faire l'objet de stratégies de contrôle qui tiennent compte du fait qu'ils surévaluent les conséquences à court terme.

Les stratégies les plus efficaces se distinguent par le fait qu'elles agissent non pas seulement sur la disponibilité des options et leur degré de désirabilité, *mais plus spécifiquement sur la distribution temporelle de leurs conséquences*.

En effet, il est communément admis que les stratégies de contrôle de soi peuvent jouer sur deux tableaux.

Sur le premier, l'agent modifiera l'étendue des options disponibles. Cela correspond à la stratégie d'Ulysse. Un alcoolique videra son stock de bouteilles dans les toilettes, un toxicomane ira en cure loin des centres de distribution de drogue, un étudiant ira faire ses travaux à la bibliothèque, où il n'y a pas de télévisions ou de consoles de jeux, un retardataire matinal arrachera le bouton « snooze » de son cadran, un conjoint adultère évitera les bars et lieux de rencontre, une personne dépensière optera pour une formule de retrait automatique sur ses paies à destination d'un compte fermé, etc.

Sur le second tableau, l'agent modifiera le degré de désirabilité des options disponibles, soit en rendant certaines options plus désirables, soit en rendant d'autres options moins désirables. Par exemple, en annonçant à nos connaissances qu'on a pris la décision d'arrêter de fumer, on rend l'option de fumer – du moins en public – moins attrayante parce qu'on risque les railleries. En pariant avec quelqu'un qu'on serait capable d'accomplir une besogne dans les délais, on peut risquer de perdre une somme suffisamment importante pour nous motiver à le faire. À l'inverse, en étudiant avec des collègues de classe, on joint l'utile à l'agréable, et en décidant d'intégrer les Alcooliques Anonymes, on obtiendra des félicitations et des signes d'encouragement découlant de notre abstention de boire. Aussi, les stratégies d'évitement des stimuli délétères entrent dans cette catégorie même si elles ne consistent pas à modifier le degré de désirabilité des options. En évitant, par exemple, d'emprunter à pied un trajet qui implique de passer à proximité d'un "fast-food" dégageant des effluves titillantes, je maintiens le degré d'attrait d'un repas médiocre pour la santé suffisamment bas, ce qui revient, d'une certaine manière, à agir sur mes préférences.

Si la distribution temporelle des conséquences est le principal facteur de faiblesse de volonté, alors on doit nécessairement en tenir compte lorsqu'on décide d'agir sur l'un ou l'autre des tableaux mentionnés. Restreindre l'étendue des opportunités doit évidemment concerner en premier lieu les options (ou habitudes) qui présentent le profil hédonique qui

décroit le plus rapidement. Par contre, si un agent akratique se trouve incapable de restreindre ses options ou si le coût pour les restreindre est prohibitif (ex. : briser son téléviseur à coups de hache pour ne plus être en mesure de perdre son temps à regarder des émissions idiotes), il peut modifier la distribution des conséquences découlant des choix disponibles. Une modification adéquate doit porter sur la distribution temporelle des conséquences.

Les paris privés et publics, par exemple, ne sont efficaces que si leurs conséquences potentielles sont distribuées d'une certaine manière. Or, lorsqu'on considère les paris comme stratégie de contrôle de soi, notre attention se porte sur la qualité des récompenses et des punitions, et non sur la structure temporelle des distributions. Un agent qui parie avec lui-même qu'il réussira à accomplir une besogne dans les délais ou à s'abstenir de consommer une substance devra mettre en jeu une récompense ou une punition exceptionnelle. Par exemple, je peux parier avec moi-même que si je ramasse les feuilles mortes sur mon terrain cette fin de semaine, alors je m'autoriserai dans la soirée de dimanche une triple part de gâteau au chocolat, chose que je ne fais habituellement pas. En fait, un pari privé n'est qu'une manière hypothétique d'allouer des biens ou de distribuer des ressources dans le temps. Un agent qui se promet une triple part de gâteau n'engloutira pas nécessairement plus de parts au final. Il ne fait que distribuer les parts différemment dans le temps. Mais la chose la plus importante est que la récompense exceptionnelle doit être temporellement aussi proche que possible du moment de réussite. Il ne sert à rien de me promettre une triple part dans six mois si et seulement si je ramasse les feuilles la prochaine fin de semaine. Une telle mesure aura un impact motivationnel insignifiant. Les choses seront différentes si la récompense suit de près l'accomplissement complet de la besogne. La même chose s'applique aux paris publics. Si un pari public a un impact motivationnel important, cela ne dépend pas exclusivement de l'ampleur de l'enjeu, mais également de la position temporelle de la punition en cas de perte. Une punition relativement bénigne, comme d'amicales railleries, peut avoir un impact motivationnel suffisamment grand pour qu'un pari public fonctionne. Cela tient au fait que les railleries sont temporellement rapprochées des points de consommation de cigarette.

Au niveau social, on peut exploiter par le biais des institutions la structure dynamique des choix intertemporels pour orienter les citoyens ou limiter l'occurrence de décisions allant

de manière très marquée à l'encontre de leurs propres intérêts (Elster, 2007). Les délais obligatoires pour se marier, pour divorcer, pour contracter un prêt immobilier (loi Scrivener), pour interrompre une grossesse, pour subir une chirurgie esthétique ou pour se faire stériliser, sont des mesures institutionnelles qui encadrent les phases finales de processus qui impliquent des décisions ayant de lourdes conséquences, un impact résiduel important et difficile, voire dans certains cas impossible, à réduire une fois la décision exécutée. Les délais obligatoires dans l'achat des armes d'assaut (États-Unis) se distinguent par le fait que l'impact résiduel de telles acquisitions est plus potentiel que réel et est susceptible d'affecter davantage autrui que soi-même. En imposant un délai obligatoire, on ne limite pas seulement les comportements impulsifs, mais on limite également les comportements trop impatients parce qu'on repousse dans le temps la première vague de conséquences.

On peut justifier l'imposition de délais obligatoires par le fait que les citoyens doivent impérativement précéder certains types de décision par une réflexion et qu'ils doivent être dans un état relativement « froid » pour les prendre. C'est d'ailleurs la première raison qui vient à l'esprit quand on tente de les justifier. Mais les délais obligatoires ne favorisent pas seulement la réflexion calme, elles modifient sensiblement la structure des incitations. L'induction d'une telle modification peut ou non produire un renversement de préférences. Si une personne n'a plus envie de contracter un prêt hypothécaire parce qu'il apprend que celui-ci ne peut légalement être contracté que dix jours plus tard, alors on peut estimer qu'elle escompte l'acquisition d'une propriété immobilière de manière très prononcée – auquel cas le dispositif institutionnel lui a peut-être permis d'éviter une allocation désastreuse de ses ressources. Les obligations de délais ont souvent pour effet d'atténuer l'escompte du futur, mais cela n'implique pas que les législateurs les ont conçus dans ce but. Le mariage en est un bon exemple. Les délais prescrits évitent les mariages sur un coup de tête et incitent les personnes moins impulsives à planifier l'événement et à prévoir un moment approprié. Mais ce qui a motivé initialement l'adoption de ce genre de règle était – du moins pour le cas du Québec – que le délai permettait l'annonce publique du mariage et minimisait, par le fait même, le risque d'union de personnes déjà mariées.

La structure intertemporelle des choix et l'escompte hyperbolique jouent également un rôle dans certains comportements immoraux et peuvent faire à ce titre l'objet de modifications

institutionnelles. La tricherie et autres comportements de resquilleur (*free-riding*) sont beaucoup plus attirants si la jouissance des bénéfices ou l'évitement des coûts sont accessibles dans l'horizon du court terme. Certaines personnes sont systématiquement disposées à mentir même si elles savent que le risque de se faire prendre est très grand, voire inévitable. Mais, en mentant, elles évitent des désagréments immédiats. Aussi, les personnes sont d'autant plus enclines à se conformer à leurs principes moraux si les bénéfices de la défection sont temporellement éloignés. La promptitude à punir les tricheurs a souvent un effet correcteur plus grand que la grandeur de la punition. Du point de vue législatif, la fonction préventive des punitions ne peut être remplie de manière satisfaisante que si les procédures entamées contre les coquins sont enclenchées dans des délais assez courts suivant les délits.

Les stratégies institutionnelles de contrôle adaptées à la structure intertemporelle des choix et l'escompte hyperbolique ne visent pas seulement les individus, mais peuvent également porter sur les groupes sociaux, voire des nations entières. Jon Elster voit, par exemple, dans les éléments constitutionnels dont se dotent les États, des dispositifs contraignants pour l'avenir, analogues à la stratégie d'Ulysse et produisant des résultats issus de processus décisionnels encadrés favorisant des intérêts à long terme. Par exemple, les politiques monétaires des États ne sont pas sous la juridiction des assemblées législatives, mais de la responsabilité des banques centrales. Non seulement cela permet d'éviter que la politique monétaire soit l'objet de velléités partisanes, mais la structure de l'organisation, les connaissances de ses membres, leurs indépendances vis-à-vis des autres organes du pouvoir, et la fonction générale de l'organisation telle que constitutionnellement prescrite, favorisent la prise de décision en fonction de l'intérêt monétaire du pays à plus ou moins terme (Elster, 2007). Les ministères de l'environnement seront peut-être appelés un jour à être désenclavés de la même façon que les organes de décision monétaire pour répondre aux défis environnementaux que pose la justice intergénérationnelle. La pollution et la dilapidation des ressources naturelles découlant de l'exploitation intensive du territoire ne sont politiquement cautionnées que dans la mesure où elles sont le résultat de la poursuite d'objectifs sociaux à plus ou moins court terme et concernant les quelques générations qui en sont les tributaires actuels.

11.2 De l'esprit étendu à la volonté étendue

Les stratégies individuelles et mesures institutionnelles que les personnes et groupes mettent en place sont souvent suffisamment efficaces pour leur assurer une réussite minimale dans la poursuite d'objectifs à long terme. La poursuite de ces objectifs nécessite la plupart du temps l'accomplissement de nombreuses besognes, d'activités déplaisantes et de périodes de renoncement qui impliquent à l'occasion des expériences hédoniques négatives assez marquées. L'examen des stratégies les plus efficaces met cependant en relief le fait que les personnes mettent en place des *échafaudages* (*scaffolding*) cognitifs *et* motivationnels dans leur environnement, dont la disparition, même momentanée, engendre une accentuation du caractère hyperbolique de leur escompte du futur.

La description empirique des échafaudages cognitifs a donné lieu dans la littérature philosophique à une reconsidération de la notion d'esprit. David Chalmers (1998) et Andy Clark (1998, 2010) ont développé l'idée que la frontière de la peau est un élément non pertinent pour circonscrire les processus cognitifs. Par exemple, la confection d'une liste d'épicerie, la résolution d'une équation mathématique sur une feuille de papier, la délimitation d'un territoire à l'aide de piquets ou d'urine, le fait de tourner sur elle-même une pièce tridimensionnelle pour essayer de la faire passer par un orifice, sont des événements qui doivent être décrits comme des phases importantes de processus cognitifs qui ont lieu en partie à l'extérieur du cerveau. Dans le cadre d'une conception internaliste (classique) de l'esprit, rechercher des mots au Scrabble, par exemple, revient à former des images mentales, ou représentations plus abstraites, de mots qui commandent à la sortie (*output*) des actes d'agencements de lettres. Si le résultat n'est pas satisfaisant, de nouvelles transformations mentales sont initiées à partir de l'agencement préalablement obtenu et disponible en entrée (*input*). Ici la résolution de problème implique des séries de boucles d'actions et rétroactions entre les processus cognitifs d'une part et les agencements d'autre part. Mais dans le cadre d'une conception externaliste comme celle que proposent Chalmers et Clark, l'agencement des lettres fait partie intégrante du processus cognitif. Pour ces chercheurs, manipuler des lettres réelles n'est pas essentiellement différent que de manipuler des lettres mentales. C'est pour cette raison qu'ils se font les avocats d'un externalisme actif en philosophie de l'esprit. Cette position se distingue de l'externalisme

issu du célèbre débat portant sur les déterminants de la référence des mots et du contenu des croyances, dont Putman et Burge furent les plus grands partisans. Pour Chalmers et Clark,

> « *when I believe that water is wet and my twin believes that twin water is wet, the external features responsible for the difference in our beliefs are distal and historical, at the other end of a lengthy causal chain. Features of the present are not relevant: if I happen to be surrounded by XYZ right now (maybe I have teleported to Twin Earth), my beliefs still concern standard water, because of my history. In these cases, the relevant external features are passive. Because of their distal nature, they play no role in driving the cognitive process in the here-and-now. This is reflected by the fact that the actions performed by me and my twin are physically indistinguishable, despite our external differences.* » (1998 : 14)

L'externalisme actif de Chalmers et Clark va plus loin que simplement considérer les supports tels que les listes d'épicerie, les dictionnaires, les fichiers informatiques, les lignes peintes sur les routes ou les traces d'urine pour certains animaux, comme des manières de décharger (*offloading*) des informations dans l'environnement pour alléger la mémoire des agents. À l'aide de supports informationnels, les agents peuvent littéralement traiter les informations à l'extérieur de leur cerveau, comme lorsqu'on fait des calculs sur du papier, lorsqu'on tape à la machine ou même lorsqu'on touche des objets pour en déterminer leur profondeur spatiale. Pour Chalmers et Clark, ce genre de cas montre que

> « *the relevant external features are active, playing a crucial role in the here-and-now. Because they are coupled with the human organism, they have a direct impact on the organism and on its behavior. In these cases, the relevant parts of the world are in the loop, not dangling at the other end of a long causal chain. Concentrating on this sort of coupling leads us to an active externalism, as opposed to the passive externalism of Putnam and Burge.* »[123]

Le traitement des informations par l'esprit peut avoir lieu en partie à l'extérieur de « son » cerveau. La division du travail intellectuel en est sans dote l'exemple le plus évident. Demander l'avis d'un expert pour solutionner un problème revient à enclencher un processus de traitement de l'information que nos cerveaux non experts sont incapables d'assumer seuls. Mais les cerveaux ne sont pas toujours réduits à fournir des informations à des systèmes externes qu'ils ne contrôlent pas dans le but de récolter ensuite les données utiles. Ils peuvent jouer un rôle actif dans le traitement externe de l'information et même en être les principaux agents. C'est ce qui se passe lorsqu'on manipule des lettres au Scrabble, lorsqu'on effectue une division en s'aidant d'un crayon et d'une feuille, lorsqu'on élabore

[123] *Ibid.*

une carte et qu'on s'en sert pour se diriger, lorsqu'on classe des items en fonction de catégories pour les retrouver plus facilement ou lorsqu'on utilise des procédés mnémotechniques comme l'inclusion de données dans une séquence complexe, mais plus facile à retenir[124].

Bien qu'ils ne développent pas cet aspect de l'externalisme actif, Chalmers et Clark suggèrent que l'esprit étendu doit comprendre un chapitre sur la *volonté étendue*. Si cela est effectivement le cas, il y aurait non seulement des échafaudages que les agents mettent place pour solutionner des problèmes cognitifs, mais également des échafaudages pour solutionner des problèmes motivationnels. Un emploi du temps, par exemple, peut être élaboré dans un agenda, et orienter notre attention sur les tâches que nous avons à faire de manière séquentielle. Consulter notre agenda est une manière active de gérer notre attention et pas seulement une manière de se rappeler quelque chose ou de pallier un problème de mémoire de travail. Bien sûr, on peut dire qu'on a géré notre attention *à l'aide* d'un agenda – auquel cas la consultation du support est conçue comme quelque chose d'externe aux processus de gestion de notre motivation. Mais on peut également dire que la consultation de l'agenda *est* la gestion directe de notre motivation. De la même manière, on peut dire que nous avons écrit un courriel à l'aide d'un ordinateur, ou ne pas mentionner cet aspect et dire plutôt que nous avons tout simplement écrit un courriel. Clark a décrit comment les dispositifs linguistiques nous aident à gérer notre attention et celle d'autrui, et nombre d'autres aspects qui permettent aux agents disposant d'une culture linguistique – même minimale – d'étendre leur esprit à l'extérieur de leur cerveau[125].

Les philosophes Joel Anderson et Joseph Heath ont exploité l'idée de volonté étendue dans un article éclairant. Pour ces chercheurs, il est incorrect de croire que les problèmes motivationnels résident entièrement dans la tête des agents. Si l'on enlève à des agents l'accès à des dispositifs externes de traitement de l'information, ils verront une réduction importante de leurs capacités cognitives. De la même manière, si on leur enlève l'accès à

[124] Par exemple, on apprend aux enfants du primaire à se rappeler des conjonctions de coordination à l'aide de la séquence « mais où et donc car-ni-or ». Non seulement la séquence est plus facile à retenir que les conjonctions prises individuellement, mais il est également facile d'inférer les conjonctions à partir de la séquence (on ne compte pas le « et » et l'on extrait les conjonctions à partir de leurs propriétés sonores).

[125] Voir *Supersizing The Mind* (§3). Le langage permet en outre d'encoder des informations pour faciliter leur traitement ou l'étendre dans des proportions autrement impossibles à obtenir. Par exemple, on voit mal comment les capacités d'évaluations des quantités auraient pu se développer sans l'utilisation de marqueurs numériques, même minimaux comme des traits gravés sur une paroi ou des objets qui jouent un rôle dans un système d'appareillage biunivoque.

des échafaudages motivationnels, les agents se montreront davantage impatients ou procrastinateurs.

Par exemple, il est en général préférable d'être abonné à un gymnase pour effectuer des exercices de manières régulières. Les stimuli qui sont présents dans ce genre d'environnement nourrissent, voire exacerbent, l'envie de fournir un effort physique important sur une période temporelle plus longue que si nous restons à la maison. Même si nous disposons de tout l'équipement nécessaire, observer des personnes fournir des efforts, ainsi que le résultat que l'entraînement physique a eu sur eux, constituent souvent le principal incitatif qui nous pousse *in situ* à continuer en dépit du fait que les résultats espérés ne commenceront à se faire sentir que dans plusieurs mois. D'ailleurs, ce n'est pas seulement la présence de certains stimuli qui fait la différence pour la motivation, mais l'absence de certains autres. Bien que beaucoup de clubs de mise en forme possèdent des téléviseurs, rares sont ceux qui possèdent en plus des canapés confortables et offrent des boissons alcoolisées. Dans ce genre de scénario, c'est la gestion des stimuli qui distingue souvent la réussite de l'échec. Cette gestion peut être très complexe suivant les cas. Certains agents doivent en plus préparer leur décision de se rendre au club en planifiant l'heure de départ à un moment où les sources de distraction ont un impact minimal (ex. : tout de suite après le travail et non après le souper pendant la plage télévisuelle la plus intéressante).

Cela n'implique pas que les mécanismes *internes* de contrôle de soi responsables de la force de volonté ne jouent aucun rôle dans l'exécution de plans à plus ou moins long terme. Ils jouent en fait un rôle essentiel. Un agent rationnel doit être minimalement capable de repousser dans le futur, voire renoncer à des satisfactions, sans quoi il ne pourrait même pas faire usage de stratégies comme celle d'Ulysse. En effet, mettre en place de telles stratégies implique un minimum d'efforts et de désagréments (ex. : mes chers marins, attachez-moi au mat!). Par contre, comme le soutiennent Anderson et Heath,

> « *If the demands of individuals for processing decisions and resisting various temptations is on the rise in increasingly complex societies, these individual capacities might easily become overtaxed – if pure willpower were all that people could rely on. It is, however, unclear to what extent the average person actually relies on individual self-discipline to avoid or to limit procrastination. Our suspicion is that its role is greatly exaggerated. Much of the time, what looks like sheer willpower is the result of more-or-less well-orchestrated attempts by individuals to*

> *arrange their lives in such a way as to economize on willpower, by avoiding situations that call for its exercise.* » (2010 : 239)

En acceptant l'idée que les agents peuvent étendre leur volonté à l'extérieur de leur cerveau, on doit, je pense, accepter également l'idée générale que les problèmes motivationnels comme la faiblesse de volonté ne résident pas entièrement dans la tête des agents. Une des difficultés académiques les plus importantes pour bon nombre d'étudiants qui entrent au cégep est due à un problème de procrastination systématique exacerbé justement par le fait que leur milieu de vie (ex. : intégration du premier appartement) a radicalement changé et que les signaux qui déclenchaient auparavant des exécutions de plan (ex. : les attentes des parents, les couvre-feux, les risques de réprimandes, etc.) ne sont plus présents et que les stimuli délétères abondent (ex. : nouvelles occasions de sortie entre amis, disponibilité étendue du téléviseur, etc.). Cela prend un certain temps avant que les étudiants arrivent à intégrer un mode de gestion plus efficace de leur volonté fondée sur des échafaudages (ex. : étudier à la bibliothèque entre les cours avec des collègues de classe) – et certains n'y parviennent jamais. Cela milite en faveur de l'idée que la faiblesse de volonté est – au moins dans certains cas – un problème motivationnel situé en partie à l'extérieur de la tête des agents. Des personnes continentes qui transitent dans des environnements très différents peuvent devenir, du jour au lendemain, incontinentes. Elles ne peuvent plus, dans certains cas, s'appuyer sur des structures d'incitations qui atténuent grandement leur dévaluation du futur. Au demeurant, avec des mécanismes internes assez peu robustes, une personne peut utiliser des stratégies de contrôle de soi externes très efficaces justement parce qu'elles s'avèrent ménager sa volonté, de la même manière que de faire une liste d'épicerie permet de ménager sa mémoire. Dans les deux cas, les agents étendent des dispositifs à l'extérieur de leur cerveau, soit pour faciliter le traitement informationnel, soit pour faciliter la prise de décision conforme au principe d'égale considération de toutes les phases temporelles de l'agent.

Ignorer le rôle que jouent les échafaudages motivationnels dans le contrôle de soi nous amène, selon Anderson et Heath, à adopter une conception mentaliste, individualiste et volontarisme du choix rationnel, alors que l'autodiscipline est en fait une propriété contextuelle née de l'interaction entre la situation et la personnalité de l'agent.

D'ailleurs, l'accès à des échafaudages non volontaires est un défi politique important. Les règlements concernant la consommation d'alcool, comme les points de vente (limités à des succursales contrôlées directement par l'État), les taxes (surtaxe des produits bas de gamme), les modes de vente (étalages et choix libres ou distribution par commande) ou les « last-call » dans les brasseries restreignent la consommation en allégeant les dispositifs internes de contrôle de soi. Or, on ne pourrait pas obtenir les mêmes résultats si l'on abrogeait ces règlements et offrait à la place des stratégies d'intervention psychologique directe utilisées sur une base volontaire (ex. : essayer d'éviter de diriger son attention sur les points de vente de produits alcoolisés, prendre des douches glacées pour renforcer sa volonté, se rappeler dans les moments cruciaux les inconvénients liés à la surconsommation, etc.). Les agents peuvent modifier volontairement leurs environnements physiques et sociaux de manière à échafauder adéquatement leur volonté. Mais cela demande minimalement un diagnostic de ses propres problèmes de faiblesse de volonté, ainsi que de savoir comment élaborer des échafaudages efficaces et identifier des stratégies suffisamment indirectes pour qu'elles soient à l'intérieur de la portée de notre motivation.

Les inégalités dans l'accès aux échafaudages motivationnels posent également un défi politique. La concentration de certains stimuli délétères (ex. machine à sous, "fast-foods", coût prohibitif des aliments sains, délabrement des infrastructures, prêteurs sur gages, boutiques « insta-meuble », etc.) dans des quartiers défavorisés peuvent faire l'objet de mesures ciblées. Bien que de telles mesures doivent être évaluées en fonction de leur efficience, elles doivent également se justifier au niveau abstrait des principes, sans quoi elles risquent d'être des mesures plus ou moins arbitraires, à saveur paternaliste.

Une des pistes serait à mon avis d'utiliser la théorie des *capabilités* pour justifier l'imposition d'échafaudages motivationnels. Initialement proposée par l'économiste Amartya Sen, la théorie des capabilités vise à pallier les insuffisances de la théorie rawlsienne fondée sur la notion de biens premiers (ex. : sécurité, opportunités, santé, liberté, intelligence, revenus, les droits, les bases sociales du respect de soi-même, etc.). Pour Sen, une distribution maximale – mais compatible avec des contraintes égalitaires minimales – de biens premiers dans la société n'est pas suffisante pour assurer aux individus des conditions favorables pour la réalisation de leurs propres projets de vie (Sen : 1993). Les individus doivent être en plus capables d'en faire usage. La création

d'opportunités – pour ne prendre que ce bien premier – pour tous ne s'accompagne pas de la production d'*habiletés* à les exploiter. Certains individus seront en mesure de le faire, d'autres non. Par exemple, l'éducation publique postsecondaire est en principe accessible à tous, mais il s'avère qu'un nombre très important d'individus issus des classes défavorisées n'y ont pas accès ou se trouvent cognitivement incapables de fournir un rendement suffisant pour pouvoir passer d'une étape à l'autre, et obtenir finalement un diplôme.

Bien qu'il s'agisse d'un concept technique, il est assez naturel de définir les capabilités comme des habiletés à exploiter des ressources – au sens large du terme – disponibles. Si un agent est capable de nager *hic et nunc*, alors il possède l'habileté nécessaire pour le faire et a accès à un plan d'eau à proximité. Mais rien ne s'oppose à inclure également des facteurs motivationnels dans l'équation. Des individus peuvent se révéler incapables d'arrêter de fumer, bien qu'ils disposent des habiletés minimales pour le faire – il n'est pas difficile d'écraser une bonne fois pour toutes! En mettant en place des échafaudages motivationnels adéquats, on rend des personnes akratiques capables d'améliorer significativement leurs sorts.

Agir sur les motivations des agents par le biais de dispositifs institutionnels est politiquement et éthiquement délicat. Le spectre du paternalisme d'État peut toujours être brandi dans les débats publics sur la question. Mais il ne fait pas de doute que l'émergence, sans précédent dans l'histoire humaine, d'une quantité aussi vaste et toujours plus grande d'opportunités et de sources de distractions pose de nouveaux défis politiques. La faiblesse de volonté est l'un de ceux-là.

Au-delà des enjeux théoriques que posent les solutions institutionnelles aux problèmes de la faiblesse de volonté, envisager socialement de telles solutions implique un changement profond d'attitude envers nos désirs. Non seulement Héraclite avait raison de dire que les hommes ne s'en tireraient pas mieux s'ils obtenaient systématiquement tout ce qu'ils veulent, mais nous devons considérer que la réussite personnelle n'implique pas toujours la satisfaction des désirs, mais quelques fois (et plus souvent qu'on le pense) leur frustration.

Bibliographie

Ainslie, George & Haslam, Nick. (1992) « Hyperbolic Discounting » in *Choice over Time*, Loewenstein, G. & Elster, Jon (dir.), New York, Russell Sage, 57-92.

Ainslie, George & Haslam, Nick. (1992) « Self-Control » in *Choice over Time*, Loewenstein, G. & Elster, Jon (dir.), New York, Russell Sage, 177-211.

Ainslie, George. (1974) « Impulse Control in Pigeons », *Journal of the Experimental Analysis of Behavior*, vol. 21, 485-9.

Ainslie, George. (2001) *Breakdown of Will*, Cambridge, Cambridge University Press.

Akerlof, G. A. (1991) « Procrastination and obedience », *American Economic Review,* vol. 81:2, 1–19.

Anderson, Joel & Heath, Joseph. (2010) « Procrastination and the Extended Will » in *The thief of Time*, Andeou C. & White, M., New-York, Oxford University Press, 233-52.

Anne-Marie Kalis, Andreas Mojzisch, T. Sophie Schweizer, Stefan Kaiser. (2008) « Weakness of will, akrasia, and the neuropsychiatry of decision making : An interdisciplinary perspective », *Cognitive, Affective, & Behavioral Neuroscience*, vol. 8:4, 402-17.

Ariely, Dan & Carmon, Ziv. (2003) « Summary Assessment of Experiences : The Whole Is Different from the Sum of Its Parts », in *Time and Decision*, Loewenstein, G. & Read, D. & Baumeister, R. (dir.), New York, Russel Sage, pp. 323-49.

Ariely, Dan. (2010a) *Predictably Irrational, Revised and Expanded Edition: The Hidden Forces That Shape Our Decisions*, New York, Harper Perennial.

Ariely, Dan. (2010b) *The Upside of Rationality : The Unexpected Benefits of Defying Logic at Work and at Home*, New York, Harper-Collins.

Aristote. (1991) *Éthique à Eudème*, (trad. Décarie), Paris, Vrin.

Aristote. (1999) *De l'âme*, (trad. Bodéüs), Paris, Garnier-Flammarion.

Aristote. (2004) *Éthique à Nicomaque* (trad. Bodéüs), Paris, Flammarion.

Aristote. (2004) *Politique* (trad. Bodéüs), Paris, Flammarion.

Aubry, Romeyer-Dherbey & Gwenaëlle, Gilbert. (2002) *L'excellence de la vie : sur l'Éthique à Nicomaque et l'Éthique à Eudème d'Aristote*, Paris, Vrin.

Audi, Robert. (1990) « Weakness of Will and Rational Action », *Australasian Journal of Philosophy*, vol. 68:3, 270-81.

Audi, Robert. (1993) *Action, Intention and Reason*, New-York, Cornell University Press.

Austin, J.L. (1961) *A Plea for Excuse*, Philosophical Papers, Oxford, Clarendon Press.

Axelrod, Robert & Cohen, Michael. (1999) *Harnessing Complexity : Organizational Implication of a Scientific Frontier*, New York, The Free Press.

Axelrod, Robert. (1984) *The evolution of cooperation*, New York, Basic Books.

Bachelard, Gaston. (1938/1980) *La formation de l'esprit scientifique*, Paris, Vrin.

Baillargeon, Robert. (1987) « Object permanence in 3 ½-and-4 ½-month-old infants », *Developmental Psychology*, 23, 655-664.

Baron-Cohen, S. (1999) « La cécité mentale dans l'autisme : Théorie de l'esprit, ancêtre, précurseurs et dysfonctionnements », Revue des Presses Universitaires de France, n° 3:1, 285-293.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985) « Does the autistic child have a" theory of mind? », Cognition, vol. 21, 37-46.

Batson, C.D. (1991) *The altruism question : Toward a social-psychological answer*, Hillsdale, Erlbaum.

Batson, C.D. (1998) « Altruism and prosocial behavior » in *The handbook of social psychology*, Gilbert, D. & Fiske, S.T. (dir.), vol.2, 262-316.

Baumeister, Roy & Heatherton, Todd & Tice, Diane. (1994) *Losing Control : How and Why People Fail at Self-regulation*, San Diego, Academic Press.

Bechara, A. & Damasio, H. & Damasio, A.R. (2000) « Emotion, decision making and the orbitofrontal cortex », *Cerebral Cortex*, vol. 10, 295-307.

Becker, Gary. (1976) The Economic Approch to Human Behavior, Chicago, University of Chicago Press.

Bentam, Jeremy. (1789/1988) *Introduction to the Principles of Morals and Legislation, New York, Prometheus Books.*

Berridge, K.C. & Robinson, T.E. (2003) « Parsing reward », *Trends in Neurosciences,* 26:9, 507-13.

Berridge, K.C. & Robinson, T.E. (2008) «The incentive sensitization theory of addiction: some current issues », *Philosophical Transactions of the Royal Society B, 363*, 3137-46.

Berridge, K.C. & Winkielman, P. (2003) « What is an unconscious emotion? » *Cognition & Emotion,* 17, 181-211.

Berridge, K.C. & Robinson, T.E. (1995) « The mind of an addicted brain: neural sensitization of wanting versus liking », *Current Directions in Psychological Science,* vol. 4, 71–6.

Boudon, Raymond. (1999) *Le sens des valeurs*, Paris, PUF.

Boudon, Raymond. (2003) *Raison, bonnes raisons*, Paris, PUF.

Bratman, Michael. (1997) « What is the Accordion Effect? », *The Journal of Ethics*, Vol. 10:1-2, 5-19.

Bratman, Michel. (1987) *Intention, Plans, and Practical Reason*, Cambridge (Mas.), Harvard University Press.

Bratman, Michel. (1997) « Pour une théorie modeste de l'action planifiée : réponse à Gauthier et Dupuy », in *Les limites de la rationalité* (t.1), Dupuy, J.-P. & Livet, P. (édit.), Paris, La Découverte, 75-87.

Bratman, Michel. (2001) « Two Faces of Intention », in *The Philosophy of Action*, Mele, A. (edit.), Oxford, Oxford University Press.

Broome, John. (1991) *Weighing Goods: Equality, Uncertainty, and Time*, Oxford, Basil Blackwell.

Broome, John. (1992) « Discounting the Future », *Philosophy & Public Affairs*, vol. 20, 128-56.

Burka, Jane B. & Yuen, Lenora M. (1983) *Procrastination: Why You Do It, What To Do About It Now*, Da Capo, Cambridge.

Byrne, Ruth. (2005) *Rational Imagination*, Cambridge (Ma.), MIT Press,.

Chalmer, David & Clark, Andy. (1998) « Extended Mind », *Analysis*, vol. 58, 10-23.

CHUNG, Shin-Ho & HERRNSTEIN, Richard. (1967) « Choice and delay of reinforcement », *Journal of the Experimental Analysis of Behavior*, vol. 10, 67-74.

Cialdini, R. & Schaller, M. & Houlihan, D. & Arps, K. Fultz, J. & Beaman, A. (1987) « Empathy based helping : Is it selflessly or selfishly motivated? », *Journal of Personality and Social Psychology*, vol. 52, 749-58.

Clark, Andy. (2010) *Supersizing the Mind*, Oxford, Oxford University Press.

Coombs, C. H. & Avrunin, G. (1977) « Good things satiate and bad things escalate : A Theorem of Single-Peaked Preference Fonctions in One Dimension », *Journal of mathematic psychology*, 16, 261-266

Damasio, Antonio. (1995) *L'Erreur de Descartes : La raison des émotions*, Paris, Odile Jacob.

Damasio, Antonio. (2010) *L'autre moi-même*, Paris, Odile Jacob.

Darwin, Charles. (1887/2002) Autobiography, New York, Penguin Classics.

Daston, L.J. (1988) *Classical probability in the Enlightenment*, Princeton, Princeton University Press.

Davidson, Donald. (1970) « Comment la faiblesse de volonté est-elle possible? » in *Actions et événements*, Paris, PUF, 35-65.

Davidson, Donald. (1971) « L'agir » in *Actions et événements*, Paris, PUF, 67-91.

Davidson, Donald. (1978) « Avoir une intention » in *Actions et événements*, Paris, PUF, 119-45.

Davidson, Donald. (1985) « Incoherence and Irrationality », *Dialectica*, vol. 39:4, 345-54.

Davidson, Donald. (1991) *Paradoxe de l'irrationalité*, Paris, L'éclat.

Davis, Lawrence. (1979) *Theory of action*, New York, Prentice-Hall.

Dawes, Robyn W. (1979) « The Robust Beauty of Improper linear models in Decision Making », *Psychological Bulletin*, vol. 81, 95-106.

Dawes, Robyn. (1990) « The Potential Nonfalsity of the False Consensus Effect », in *Insights Decision Making : A Tribute to Hillel*, Einhorn, J. & Hogarth, R. M. (dir.), University of Chicago Press, 179-199.

Dawkins, Richard. (1976). *The Selfish Gene*, New York, Oxford University Press.

De Sousa, Ronald. (1987) *The Rationality of Emotion*, Cambridge (Mas.), MIT Press.

De Sousa, Ronald. (2000) « Deux maximes de la rationalité émotive » in *Emotions et Rationalité*, Engehrn, E. & Baertschi, B. Verlag, Lausanne, Paul Haupt, 15-32.

De Sousa, Ronald. (2004) *Évolution et rationalité*, Paris, PUF.

De Waal, Franz. (2009) *The Age of Empathy: Nature's Lessons for a Kinder Society*, New York, Emblem Editions.

Dennett, Daniel. (1996) *Kinds of Minds*, New York, Brockman.

Devin, Henry. « Aristotle on Pleasure and the Worst Form of Akrasia », *Ethical Theory and Moral Practice* (2002), 5:3, 255-270.

Dodd, Dylan. (2007) « Weakness of Will as Intention-Violation », *European Journal of Philosophy*, vol. 17:1, 45-59.

Dutton, D. & Aron, A. (1974) « Some evidence for heightened sexual attraction under conditions of high anxiety », *Journal of Personality and Social Psychology*, vol. 30, 510-17.

Elster, J. & Loewenstein, G. (1992) « Utility from Memory and Anticipation » in *Choice over Time*, Loewenstein, G. & Elster, J. (dir.), New York, Russell Sage, 213-34.

Elster, Jon. (1979/1986) *Le laboureur et ses enfants*, Paris, De Minuit.

Elster, Jon. (1999) *Alchemy of Mind : Rationality and the Emotion*, Cambridge (Mas.), Cambridge University Press.

Elster, Jon. (2000) *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*, Cambridge (Mas.), Cambridge University Press.

Elster, Jon. (2007) *Agir contre soi*, Paris, Odile Jacob.

Euripide. (2002) *Médée*, Paris, Jai Lu.

Euwe, Max. (1969) *Les échecs : jugement et plan*, Paris, Payot.

Fabienne Pironet & Christine Tappolet. (2003) « Faiblesse de la Raison ou Faiblesse de Volonté : Peut-on Choisir? » *Dialogue,* vol. 42:04.627-44.

Feinberg, Joel. (1970) *Doing and Deserving*, Princeton, Princeton University Press.

Fischman, M.W. & Foltin, R.W. (1992) « Self-administration of cocaine by humans : A laboratory perspective » in *Cocaine : Scientific and Social Dimensions*, Bock, G.R. & Whelan, J. (dir.), vol. 166, Chichester, Wiley, § 10.

Fisette, Denis & Poirier, Pierre. (2000) *Philosophie de l'esprit : État des lieux*, Paris, Vrin.

Fisher I. (1918) « Is 'utility' the most suitable term for the concept it is used to denote? » A. N. Page (Ed.), *Utility theory: A book of readings* (49–51). New York: Wiley (Reprinted from *American Economic Review*, 8, 335–37).

Frankfurt, Harry G. (1978/2001) « The Problem of Action », in *The Philosophy of Action*, Mele, A. (dir.), Oxford, Oxford University Press, 42-52.

Frankin, Benjamin. (1706-1790/1956) *Mr. Franklin: A Selection from His Personal Letters*, Whitfield J. Bell Jr. & Leonard W. Labaree (edit.), New Haven, Yale University Press.

Frederick, Shane. (2006) « Valuing future life and future lives: A framework for understanding discounting », *Journal of Economic Psychology*. Vol. 27, 667-80.

Frederick, Shane & Loewenstein, George & O'Donoghue, Ted. (2003) « Time Discounting and Time Preference : A Critical Review » in *Time and Decision*, Loewenstein, G. & Read, D. & Baumeister, R. (dir.), New York, Russell Sage, 13-88.

Frederick, Shane & Loewenstein. *(2008)* « Conflicting Motives in evaluation of sequences », *Springer Science + Business Media, LLC.* 10.1007.

Frederick, Shane. (2003) « Time Preference and Personal Identity » in *Time and Decision*, Loewenstein, G. & Read, D. & Baumeister, R. (dir.), New York, Russell Sage, 89-114.

Gauthier, David. (1997) « Intention et délibération » in *Les limites de la rationalité* (t.1), Dupuy, J.-P. & Livet, P. (édit.), Paris, La Découverte, 59-75.

Geach, P. T. (1965) « Assertion », *Philosophical Review*, vol. 74, 449-65.

Gert, Bernard. (1988) *Morality : A New Justification of the Moral Rules*, Oxford, Oxford Press.

Gettier, Edmund. (1963) « Is Justified True Belief Knowledge? », *Analysis*, 23, 121-123.

Gigerenzer, G., Todd, P. & the ABC Research Group. (1999) *Simple Heuristics That Make Us Smart*, New-York, Oxford University Press.

Gigerenzer, Gerd. (2007) *Gut Feeling : The Intelligence of The Inconscious*, New-York, Viking Adult.

Gigerenzer, Gerd. (2008) *Rationality for Mortals*, Oxford, Oxford University Press.

Gilbert, Daniel & Wilson, Timothy. (2006) « Miswanting : Some Problem in the Forecasting of Future Affective State » in *The Construction of Preference*, Lichtenstein, S. & Slovic, P. (dir.), New York, Cambridge University Press, 550-63.

Gilbert, Daniel Todd. (2006) *Strumbling on Happiness*, New York, Vintage Books.

Goldman, Alvin. (1970) *A Theory of Human Action*, New York, Prentice-Hall.

Golstein, D. G. & Gigerenzer, G. & Hogarth, R.M. & Kacelnik, A. & Kareev, Y. & Klein, G. & al. (2002) « Group report : Why and When do simple heuristics work? » in *Bounded Rationality : The Adaptative Toolbox*, Cambridge (Ma.), MIT Press, 173-90.

Hare, Richard. (1952) *The Language of Moral*, Oxford, Oxford University Press.

Hare, Richard. (1963) *Freedom and Reason*, Oxford, Clarendon Press.

Hastie, Reid & M. Dawes, Robyn. (2010) *Rational Choice in an Uncertain World : the psychology of judgment and decision making, Londres*, *Sage.*

Heath, Joseph. (2008) *Following the Rules: Practical Reasoning and Deontic Constraint*, New-York, Oxford University Press.

Heath, Joseph. (2009) Filty Lucre : Economics for People who Hate Capitalism, Toronto, Harper-Collins.

Herrnstein, Richard. (1961) « Relative and absolute strength of response as a function of frequency of reinforcement », *Journal of the Experimental Analysis of Behavior*, vol. 4, 267–72.

Holton, Richard. (1999) « Intention and Weakness of Will », *Journal of Philosophy*, vol. 96, 241-62.

Holton, Richard. (2003) « How is Stength of Will Possible? » in *Weakness of Will and Practical Irrationality*, Oxford, Clarendon Press, 39-67.

Holton, Richard. (2006) « The act of Choice », *The Philosophers'Imprint*, vol.6:3, 1-15.

Holton, Richard. (2009) *Willing, Wanting, Waiting*, New York, Oxford University Press.

Houthakker, H. S., (1950), « Revealed preference and the utility function », *Economica* (New Series), vol. 17, 159-74.

James, William. (1884) « What is a emotion? », *Mind*, vol. 9, 188-205.

James, William. (1890/2007). *The principles of psychology*. New York, Cosimo.

Johnson-Laird, Philip. (2009) *How we Reason*, New York, Oxford University Press.

Kacelnik, Alex. (2003) « The Evolution of Patience » in *Time and Decision*, Loewenstein, G. & Read, D. & Baumeister, R. (dir.), New York, Russell Sage, 115-38.

Kagan, Shelly. (1989) *The limits of Morality*, Oxford, Oxford Clarendon Press.

Kahneman, Daniel. (2003) « A perspective on judgement and choice » *American Psychologist*. Vol. 58, 697-720.

Kahneman, Daniel. (2006) « New Challenges to the Rationality Assumption » in *The Construction of Preference*, Lichtenstein, S. & Slovic, P. (dir.), New York, Cambridge University Press, 487-503.

Kassam, K. S. & Gilbert, D. T. & Boston, A. & Wilson T. D. (2008) « Future anhedonia and time discounting », *Journal of Experimental Social Psychology*, vol. 44, 1533-7.

Kavka, Gregory. (1983) «The Toxin Puzzle », *Analysis*, 43:1, 33-36.

Keith E. Stanovich. (2004) « Balance in Psychological Research: The Dual Process Perspective », *Behavioral and Brain Sciences*, vol. 27:3, 357-358.

Keynes, J. M. (1930/1971) « Perspectives économiques pour nos petits enfants » in Essais sur la monnaie et l'économie : les cris de Cassandre, Paris, Payot, § 9.

Kihlstrom, J.F. (1999) « The psychological unconscious » in *Handbook of personality : Theory and research*, Pervin, L.A. & John, O.P. (dir.) (2nd ed.) New York, Guilford Press, 424–42.

Kim, Jaegwon. *(1976)* « Events as Property Exemplifications » in *Action Theory,* Brand, M. & Walton, D. (dir.), Dordrecht, Reidel, 159–77.

Klein, Gary. (2001) « The fiction of Optimisation » in *Bounded rationality : The adaptive toolbox*, Gigerenzer, G. & Selter, R. (dir.), Cambridge (Mas.), MIT Press, .

Korsgaard, Christine. *(*1986) « Skepticism about Practical Reason » *Journal of Philosophy*, vol. 83, 5-25.

Kosslyn, S.M. (1980) *Image and Mind*, Cambridge (Ma.), Harvard University Press.

Lamb, R.J & Preston, K.L & Schindler, C.W. & Davis, F. & Kartz, J.L. & Henningfield, J.E. & Golberg, S.R. (1991) « The reinforcing and subjective effects of morphine in

post-addicts : A dose-response study », *Journal of Pharmacology and Experimental Therapeutics*, vol. 259, 1165-73.

Ledoux, Joseph. (1998) *The emotional brain : the mysterious underpinnings of emotional life*, New York, Simon & Schuster.

Liberman, N., Trope, Y., Macrae, S., & Sherman, S. J. (2007) « The effect of level of construal on the temporal distance of activity enactment », Journal of Experimental Social Psychology, 43, 143-9 .

Liberman, Nira & Trope, Yaacov. (2003) « Construal Level Theory of Intertemporal Judgment and Decision » in *Time and Decision*, Loewenstein, G. & Read, D. & Baumeister, R. (dir.), New York, Russell Sage, 245-276.

Loewenstein, George & Ariely, Dan. (2006) « The Heat of The Moment : The Effect of Sexual Arousal on Sexual Decision Making », *Journal of Behavioral Decision*, vol. 19, 87-8.

Loewenstein, George & Prelec, Drazen. (1993) « Preference for Sequences of Outcome », *Psychological Review*, vol. 100:1, 91-108.

Loewenstein, George. (1988) « Frames of mind in intertemporal choice », *Management Science*, 34, 200-214.

Loewenstein, George. (2005) « Hot-cold empathy gaps and medical decision making », *Health Psychology* 24:4, 49-6.

Loewenstein, George. « Out of control : Visceral influences on behavior », *Organizational Behavior and Human Decision Processes*, vol. 65, 272-92.

Loewenstein, Goerge & Small, Deborah (2007) «The Scarecrow and the Tin Man : The vicissitudes of human sympathy and caring », *Review of General Psychology*, 11, 112-26.

Martignon, L. & Hoffrage, U. (1999) « Why does one-reason decision making work? : A case study in ecological rationality » in *Simple Heuristics That Make Us Smart*, Gigerenzer, Gerd & Todd, Peter & the ABC Research Group (dir.), New-York, Oxford University Press, 119-40.

McIntyre, Alasdair. (1990) « Is Akratic Action Always Irrational? », in *Identity, Character, and Morality*, O. Flanagan and A. Rorty (eds.), Cambridge, Massachusetts, MIT Press, 379-400.

McIntyre, Alasdair. (2008) « Conflicts of Desire » in *Weakness of Will from Plato to the Present*, Hoffmann, T. (dir.), Washington, The Catholic University of America Press, 276-91.

McLeod, P. & Dienes, Z. (1996) « Do fielders know where to go catch the ball or only how to get there? », *Journal of Experimental Psychology : Human Perception and Performance*, vol. 22, 531-543.

Mele, Alfred. (1987) *Irrationality : An Essay on Akrasia, Self-Deception, and Self-Control*, New-York, Oxford University Press.

Mele, Alfred. (1987) *Irrationality*, New York, Oxford University Press.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989) « Delay of gratification in children », *Science*, 244, 933-8.

Mischel, Walter & Shoda, Yuichi & Rodriguez, Monica. (1992) « Delay of Gratification in Children » in *Choice over Time*, Loewenstein, G. & Elster, Jon (dir.), New York, Russell Sage, 147-66.

Murphy, J.L. & Zajonc, R.B. (1991) « Affect, cognition, and awarness Affective priming with optimal and suboptimal stimulus exposures », *Journal of Personality and Social Psychology,* 60, 181-92.

Nagel, Ernest. (1977) « Goal-directed Processes in Biology », *Journal of Philosophy*, vol. 74, 271ff.

Nathanson, Stephen. (1994) Ideal of Rationality : A Defense, Within Reason, Chicago, Open Court.

Newby-Clark, I.R. & Ross, M. (2003) « Conceiving the Past and the Future », *Personality and Social Psychology Bulletin*, 29, 807-818.

Newman, Von J. & Morgenstern, O. (1953) *Theory of games and economic behavior*, New York, Wiley.

Nietzsche, Friedrich. (1886/1987) Par-delà le bien et le mal, Paris, Gallimard.

Nisbett, Richard E. & Wilson, Timothy D. (1977) « Telling More Than We Can Know : Verbal Reports on Mental Processes », *Psychological Review*, vol. 84:3, 231-259.

Ogien, Ruwen. (1999) *Le réalisme moral* (dir.), Paris, PUF.

Ogien, Ruwen. (2003) *Le rasoir de Kant*, Paris, L'éclat.

Ogien, Ruwen. (2007) *L'éthique aujourd'hui : Maximalistes et minimalistes*, Paris, Gallimard.

Ovide. (1992) *Les métamorphoses,* Paris, Gallimard.

Parfit, Derek. (1984) *Reason and Person*, Oxford, Clarendon Press.

Pears, David. (1984) *Motivated Irrationality*, Oxford, Oxford University Press.

Pettit, Philip. (2004) *Penser en société*, Paris, PUF.

Pettit, Philip. (2010) « *Deliberation and Decision* » in A Compagnion to the Philosophy of Action O'Connor, T. & Sandis, C. (dir.), Oxford, Wiley-Blackwell, 252-8.

Pinker, Steven. (2000) *Comment fonctionne l'esprit*, Paris, Odile Jacod.

Platon. (1950) *Œuvres complètes* (t.1&2), Paris, Gallimard.

Platon. (1993) *Ménon* (trad. Canto-Sperber), Paris, Garnier-Flammarion.

Platon. (1997) *Protagoras* (trad. Ildefonse), Paris, Flammarion.

Popper, Karl. (1979) *Objective Knowledge,* Oxford, Oxford University Press.

Prelec, Drazen & Bodner, Ronit. (2003) « Self-Signaling and Self-Control » in *Time and Decision*, Loewenstein, G. & Read, D. & Baumeister, R. (dir.), New York, Russell Sage, 277-300.

Proust, Joëlle. (2005) *La nature de la volonté*, Paris, Gallimard.

Putnam, Hilary. (2004) *The collapse of the fact/value dichotomy and other essays*, Boston, Harvard University Press.

Pychyl, T.A., Lee, J.M., Thibodeau, R., & Blunt, A. (2000) « Five days of emotion: An Experience sampling study of undergraduate student procrastination », *Journal of Social Behavior and Personality*, 15, 239-54

Quine, Willard V.O. (1951) « Two Dogmas of Empiricism », *The Philosophical Review*, vol. 60, 20-43.

Rachlin H. & Green, L. (1972) « Commitment, choice and self-control », *Journal of the Experimental Analysis of Behavior*, vol. 17:1, 15-22.

Rachlin, H. & Raineri, A. (1992) « Irrationality, Impulsiveness, and Selfishness as Discount Reversal Effects » in *Choice over Time*, Loewenstein, G. & Elster, Jon (dir.), New York, Russell Sage, 93-118.

Rachlin, H. (2006) « Notes on discounting », *Journal of the Experimental Analysis of Behavior*, 85:3, 425-35.

Rawls, John. (1973/1987) *Théorie de la justice*, Paris, Seuil.

Rizzolatti, G., Folgassi, L. & Gallese, V. (2007) « *Les neurones miroirs »*, *Pour la Science*, 44-9.,

Rorty, Amelie. (1980) « Where Does the Akratic Break Take Place? », *Australasian Journal of Philosophy*, vol. 58, 333-47.

Ross, William David. (2002) The Right and the Good, Oxford, Oxford University Press.

Ryle, Gilbert. (1949) *Concept of Mind*, Londres, Hutchinson.

Saarinen, Risto. (1994) *Weakness of Will in Mediaval Thought : From Augustine to Buridan*, New York, E.J. Brill.

Samuelson, Paul A. (1938) « A note on the pure theory of consumer's behavior », *Economica*, vol. 5:17, 61–71.

Seligman, M.E.P. & Csikszentmihalyi, M. (2000) « Positive psychology: An introduction », *American Psychologist*, vol. 55, 5-14.

Sellar, Wilfrid. (1963) « Some Reflexions on Language Games » in *Science, Perception and Reality*, Londres, Routledge & Kegan Paul.

Sen, Amatya. (1993) *Éthique et économie*, Paris, PUF.

Sen, Amatya. (2005) *Rationalité et liberté en économie*, Paris, Odile Jacob.

Shafir, Eldar& Simonson, Itamar & Tversky Amos. (2006) « Reason-Based Choice » in *The Construction of Preference*, Lichtenstein, S. & Slovic, P. (dir.), New York, Cambridge University Press, 411-33.

Shefrin, Hersh & Thaler, Richard. (1992) « Mental Accounting, Saving, and Self-Control » in *Choice over Time*, Loewenstein, G. & Elster, Jon (dir.), New York, Russell Sage, 287-330.

Shelling, Thomas. (1966) « Comments » in *Strategic interaction and conflict*, Archibald, K. (dir.), Berkeley, University of California Press, 150.

Sidgwick, Henry. (1874/2009) The Methodes of Ethics, New York, BibioBazaar.

Simon, Herbert. (1956) « Rational Choice and the Structure of Environment », *Psychological Review*, vol. 63, 129-38.

Simon, Herbert. (1982) *Models of bounded rationality*, Cambridge (Ma.), MIT Press.

Smith, Michael. « Internal Reason », *Philosophy and Phenomenological Research* (1995), vol. 55:1, 109-31.

Soman, D., Ainslie, G., Frederick, S., Li, X., Lynch, J., Moreau, P., Mitchell, A., Read, D., Sawyer, A., Trope, Y., Wertenbroch, K., Zauberman, G. (2005) « The Psychology of Intertemporal Discounting: Why are Distant Events Valued Differently from Proximal Ones? » *Marketing Letters*, 16:3/4, 347-360.

Sommer M., Sodian B., Döhnel K., Schewedtner J., Meinhardt J., Hajak G. (2010) « In psychopathic patients emotion attribution modulates activity in outcome-related brain areas » in *Psychiatry Res.*, vol. 182:2, 88-95.

Stanovich, K. E. (2004). « Balance in psychological research : The dual process perspective », *Behavioral and Brain Sciences,* vol. 27, 357-358.

Stanovich, K. E., West, R. F. & Toplak, M. E. (2008) « Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions », *Journal of Educational Psychology*, vol. 100, 930-941.

Stanovich, Keith. (1999) *Who is Rational? Studies of Individual Differences in Reasoning*, Mahwah, Erlbaum.

Steel, Pier. (2007) « The Nature of Procrastination : A Meta-Analytic and Theoretical Review of Qintessentiel Self-Regulatory Failure », *Psychological Bulletin*, Vol. 133, No. 1, 65-94

Stigler, G.J. (1961) « The economics of information », *Journal of Political Economy*, vol. 69, 213-25.

Strotz, R. H. (1955) « Myopa and Inconsistency in Dynamic Utility Maximization », in *Review of Economic Studies*, vol. 3:3, 165-80.

Sunstein Cass & Thaler, Richard. (2003) « Libertarien Paternalism Is No an Oxymoron », *University of Chicago Law Review*, vol. 70, 1159-202.

Sunstein, C. & Thaler, R. (2006) « Libertarian Paternalism Is Not an Oxymoron » in *The Construction of Preference*, Lichtenstein, S. & Slovic, P. (dir.), New York, Cambridge University Press, 689-708.

Tappolet, Christine. (2003) « Emotion and the Intelligibility of Akratic Action, in *Weakness of Will and Practical Irrationality* », Stroud, S. & Tappolet, C. (dir), Oxford, Oxford University Press, 97-120.

Tappolet, Christine. (2004) « Le prescriptivisme universel de Hare », *Recherches sur la Philosophie et le Langage*, Jean-Yves Goffi (dir.), vol. 23, 177-95

Thaler, Richard. (1985) « Mental Accounting and Consumer Choice », *Marketing Science, vol.* 4, 1999-214.

Thaler, Richard. (1981) « Some Empirical Evidence on Dynamic Inconsistency », *Economic Letters*, vol. 8, 201-207.

Thompson, Travis. (2008) « Self-Awareness: Behavior analysis and neuroscience » *The Behavior Analyst*, 31, 137-44.

Tversky, Amos. & Kahneman, Daniel. (1974) « Judgement under uncertainty : Heuristics and biases », *Science*, 185, 1124-1130.

Tversky, Amos. (1972) « Elimination by aspects: A theory of choice », *Psychological Review*, vol. 79:4, 281-99.

Vendler, Zeno. (1957) « Verbs and Times », *The Philosophical Review*, Vol. 66:2, 143-60.

Watson, Gary. (1977) « Skepticisme About Weakness of Will », *Philosophical Review* vol. 86:3, 316-39.

West, R. F., Toplak, M. E., & Stanovich, K. E. « Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions ». *Journal of Educational Psychology*, ,2008, vol. 100, 930-941.

Williams, **Bernard.** « Internal and External Reasons » in *Rational Action* (1979), Ross Harrison (dir.), Cambridge University Press, 101-13.

Winkielman, P. & Berridge, K.C. (2004) « What is an Unconscious emotion? », *Current Directions in Psychological Sciences*, 13:3, 181-212.

Ressources internet

Setiya, Kieran, « Intention », *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition),* Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/intention/>.

Structuredprocrastination.com (John Perry 1995-2008)

Procrastinus.com (Pier Steel 2010-2011)